THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXXI September, 1940 Number 6

PERMANENCE OF VOCATIONAL INTERESTS

A. C. VAN DUSEN

Psychology Department, Northwestern University, Evanston, Ill.*

INTRODUCTION

Both educational and vocational guidance make use of the individual's interests to aid him in making an adequate adjustment to his environment. If interests are conditioned by the nature of the individual's previous experiences, which in turn influence him with regard to his future experiences, then as he meets and adjusts to a changing environment, his interests may also show a change. To use interests successfully in guidance work the degree of permanence which they show becomes a problem of considerable moment. Particularly in guidance of undergraduates in college, the counselor needs to know how effectively a man's interests as a senior can be predicted in his freshman year.

The development and use of interest measurement techniques has been admirably reviewed by Fryer in his *Measurement of Interest.*[5] The one outstanding instrument for the measurement of interests which has been developed and the scale which has been used most widely in guidance work is the Vocational Interest Blank by Edward K. Strong, Jr.[8] Strong[9] has investigated the problem of permanence of interest by administering his Interest Blank to a group of two hundred twenty-three Stanford University seniors just before graduation and readministering the test to the same group five years later. His method of determining the change in interest was essentially a study of the consistency of interest score as an index. Such a treatment of change disregards the existence of a trend, an attribute of change which might well be present without influence on the correlation coefficient. Neither does his study mention the possibility of non-linear relation-

* This study was made at the University of Florida, Gainesville, Florida, 1938.

ships existing between the initial and retest scores. If non-linearity did occur in any of the vocational scales, knowledge of the nature of the function would be essential for adequate prediction.

Finch[4] and Rock[7] have used the Strong Blank in test-retest studies, but they do not supply information about the permanence of interest test scores applicable to an undergraduate college group. One investigation has been made, however, which in part is directly concerned with such a group. Burnham[1]* in investigating the stability of interest scores has administered the Vocational Interest Blank to one hundred eighty-eight Yale University students first as freshmen, then again three years later as seniors. Burnham's stability of occupational scores, just as Strong's, is expressed in the form of a correlation coefficient showing the consistency between the initial and retest scores. Stability, interpreted in the sense used in this study, that of absolute change, is not dealt with here, for the correlation may be perfect in the case where there are no changes whatever between two tests and may still be perfect if the changes are large but equal. Burnham's investigation further differs from the present in that the former also fails to report the probability or possibility of existing trends in score as well as non-linear functions in some relationships discussed.

FORMULATION OF THE PROBLEM

The purpose of this study is, first, to investigate the change in vocational interests during the college period and, second, to note, where possible, the factors influencing changes. These two aspects of the problem will be presented separately. The study is concerned only with the changes in vocational interests which occur during the college period and is based on the interest scores obtained by administering the Strong Vocational Interest Blank[8] to college students first as freshmen, then again to the same students as seniors.

The index to a change in interest used here is a change in interest score as obtained from the Strong Vocational Interest Blank. There are three attributes of a change in score with which we are concerned: (1) trend, whether positive, negative, or zero; (2) stability or total amount of change; (3) consistency exhibited between a man's score as a freshman and his score as a senior. The quantitative measures which have been used in the investigation of these attributes, if presented

* Suggested reference to this unpublished dissertation was made through correspondence with Professor Strong, and a copy was obtained only after the operations of the present study were complete.

here, will make for a better understanding of the relationships that exist between them:

(1) Trend:—M_d, Mean of the differences

$$M_d = M_s - M_f$$

(2) Stability:—RMS_d, Root mean square change

$$RMS_d = \sqrt{\frac{\Sigma d^2}{N}} = \sqrt{\sigma_d^2 + M_d^2}$$

(3) Consistency:—r_{fs}, Freshman-senior correlation

$$r_{fs} = \frac{\sigma_f + \sigma_d r_{fd}}{\sqrt{\sigma_f^2 + 2r_{fd}\sigma_f\sigma_d + \sigma_d^2}}$$

In the above equations s, f, and d denote the distribution of senior, freshman, and difference between senior and freshman scores, respectively. M denotes the mean of the given distribution; σ, the standard deviation; r, the correlation coefficient between two variables; RMS, the root mean square or quadratic mean; N, total number of cases.

M_d, being the mean of the difference scores, signifies a trend when one exists, and its sign denotes whether the trend is toward an increase or decrease in interest. RMS_d, the quadratic mean of the difference scores, is an index to the stability of score in that the quadratic mean is here a measure of how much the absolute change in score differs from zero, or point of no change. Such a measure is superior to the mean or the standard deviation of the differences since either one at a given time may be zero and high instability may still exist. r_{fs} is an index to the concomitant variation which exist between freshman and senior scores and measures the extent to which a man's interest as a senior can be predicted from a test taken his freshman year. Examination of the above formulae reveals the presence of certain degrees of interdependence between the three attributes of change in score.

Perfect consistency implies either that the standard deviation of the changes is zero or that the correlation between freshman and difference scores is one. The quadratic mean change may or may not be zero, depending upon the value of M_d or σ_d, neither of which is necessarily zero, even with perfect consistency. Inconsistency implies instability. No relationship exists directly between consistency and trend, since the correlation coefficient is independent of the mean change.

The second aspect of the problem, that of determining, where possible, the factors which influence a change in interest, has two separate parts. The first, using internal criteria, includes the determination of any relationship which may exist between the initial freshman score and a change in score using the correlation coefficient as an index (r_{fd}). If the amount of change is dependent upon the value of the initial score this fact would be shown by such an analysis. Further, the relationship existing between the changes in score in one vocation and those of each other is investigated.

The second part of this investigation involves the comparison of the upper and lower quartiles of the distribution of changes in score for each vocation with external factors which may make for or indicate change in interest.

THE INVESTIGATION

The University of Florida men who, in seeking guidance in the Bureau of Vocational Guidance and Mental Hygiene, filled out the Strong Vocational Interest Blank as freshmen at least thirty-five months prior to this study and who were still in the University constituted the subjects of this study. At the time the study was made, seventy-six such cases were found. The Blank was readministered to these men and it is upon the scores obtained from the initial and retest blanks that the investigation is based. Such a time interval permits a man to have taken the test in his freshman year and be tested again during his senior year. All of the subjects except three were seniors, law, or graduate students at the time the second blank was filled out. These three had failed to earn enough credits to be classified by the administration as of senior or equivalent status. A mean time interval of forty-one months elapsed between the administration of the freshman test and the senior test. Two-thirds of the cases were found to have test-retest intervals of thirty-five to thirty-nine months.

It is desirable to inquire into possible influences which might make this sample atypical. The subjects were all University of Florida men, but there seems to be no reason to suppose that the University of Florida should have an effect upon patterns of interest which is widely different from that of any other American university.

The fact that these individuals voluntarily sought guidance may or may not present a sampling effect on the results of this study. As the Bureau of Vocational Guidance offers its services free to any student wishing to take advantage of them, all types of students seek informa-

tion. Some have fairly definite plans concerning the vocation which they are to enter and wish merely to confirm their decision with test results and professional advice. Others, having no vocation definitely in mind, go to the Bureau for suggestion. A third group is made up of those individuals who go along with their friends to the Bureau as a matter of curiosity. Thus it is difficult to determine what effect, if any, voluntarily seeking guidance might have on the aspects of this problem.

The nature of the investigation necessitates the individual's remaining in the University long enough to enter upon his fourth year. It follows that the intelligence of the subjects is likely greater than that of the average university student since they have survived the mortality tendency of students matriculating in college. The mean percentile score on the American Council on Education Psychological Examination for this group was found to be 64.7, which is above the average for college freshmen. Further, an unpublished study by the Bureau of Vocational Guidance shows that students who seek guidance tend to be drawn from the upper deciles on this psychological test. However, it is questionable that this feature would affect permanence of interest. The fact that the individual remains in school four years or more may mean that he not only has more than average intelligence, but that there is a strong motivation present which may or may not have some influence upon the stability of his interests. The average age of the group at the freshman and senior administrations was eighteen and two-tenths and twenty-one and six-tenths years, respectively.

With the facilities available it was impractical to include in this study every vocation for which keys have been prepared. Strong, on the basis of interest,[8] has grouped the thirty-odd vocations into five categories. An individual's score on a representative vocation in each of these groups is an index to his interest score in each of the other vocations included within that group. The selection here of a representative vocation from each of the five categories was based upon two things. First, the existence of a high average intercorrelation within its own group as found by Strong, and, second, the frequency of requests for scoring among those seeking vocational guidance.[2] The following vocations were chosen as representative: Group I, Engineer; Group II, Lawyer; Group III, Teacher; Group IV, Office Clerk; Group V, Certified Public Accountant.*

* A recent announcement from the Stanford University Press concerning Strong's new *Group Scales* indicates that the vocations selected for the present study

Each freshman and senior blank was scored on the five representative vocations and the vocation which was most closely related to the vocational choice indicated by the individual. The difference score was computed for each individual in each vocation studied. The raw cores were converted into quartile (Q) scores for the use in the statistical operations. The relationship between Q scores and Strong's scores is given in Table I where $Q = \frac{S - Md}{Q.D.}$.

TABLE I.—RELATIONSHIP BETWEEN Q SCORES AND STRONG'S LETTER GRADES

STRONG'S GRADES	Q SCORE
A	−1 and greater
$B+$	−2 to −1.01
B	−3 to −2.01
$B-$	−3.5 to −3.01
C	below −3.5

The quartile scores allow comparison of scores within the same vocation on two blanks and they are also more discriminative than Strong's score system of A, B, and C. They are particularly useful, however, in cross-comparisons, that is, the comparison between an individual's scores on several different vocations. Such comparisons between vocations (although not those within a single vocation) are based upon the assumption that the median score of one vocational group represents a degree of interest equivalent to the median score of any other vocational group, and upon the further assumption that the dispersion or quartile deviations are equivalent for the various standardization groups. It is reasonable to suppose that the interest of the average lawyer in law is equivalent to the interest of the average physician in medicine. The two assumptions are implicit in Strong's use of A, B, and C scores.

Within an individual, changes in interest may differ in his chosen profession from those in a randomly selected vocation. This possibility was investigated by making a study of the vocational choices indicated on either the freshman or senior blank in addition to the five vocations mentioned above. These freshman choices and senior choices were respectively grouped together, regardless of the different vocations which they included, on the basis of the assumptions above

may well represent five of his six new scales: Engineer is representative of Strong's Group II; Lawyer, X; Teacher, V; Office Clerk, VIII; Certified Public Accountant, VII. See E. K. Strong's *Manual for Vocational Interest Blank.* Stanford University Press, 1939.

and treated just as two additional vocations, Freshman Choice and Senior Choice. Of the sixty-three individuals who expressed a choice on both the freshman and senior blanks, twenty-eight exhibited a change in their anticipated occupation while thirty-five did not. These two groups were investigated separately. A summary of the vocational choices and changes in choices is presented in Table II.

TABLE II.—FRESHMAN AND SENIOR DISTRIBUTION OF VOCATIONAL CHOICE*

Freshman Vocational Choice	Senior Vocational Choice: Certified Public Accountant	Engineer	Lawyer	Teacher	Office Clerk	Advertiser	Architect	Chemist	Doctor	Farmer	Journalist	Mathematician	Minister	Salesman	None	Total
Certified Public Accountant	4	..	..	..	1	..	..	..	..	..	..	..	..	..	..	5
Engineer	..	4	..	1	..	..	..	2	..	..	..	..	..	1	..	8
Lawyer	2	..	7	..	2	..	..	..	..	1	..	..	..	..	1	13
Teacher	..	..	1	4	..	..	..	..	..	..	..	..	..	..	..	5
Office Clerk	..	..	3	1	3	..	..	..	..	..	..	..	..	..	2	9
Advertiser	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0
Architect	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0
Chemist	1	1	..	1	..	..	..	5	..	1	1	1	1	..	..	12
Doctor	..	..	..	..	1	..	..	1	4	..	1	..	..	..	..	7
Farmer	1	..	..	..	..	..	..	..	..	1	..	..	..	..	..	2
Journalist	..	..	..	..	..	1	..	..	1	..	3	..	..	..	..	5
Mathematician	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0
Minister	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0
Salesman	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0
None	..	1	..	1	1	..	1	..	1	2	..	..	..	1	1	9
Total	8	6	11	8	8	1	1	8	6	5	5	1	1	2	4	

* Choices are classified here under the Strong vocation which most closely approximates the vocational choice.

Table II shows that Certified Public Accountant, Lawyer, and Office Clerk form an interrelated group of vocational choices in that an individual choosing one will, if he changes his choice, be likely to change to another of the group. Engineer and Chemist form another such group.

Frequency distributions of the freshman, senior, and difference quartile scores were tabulated for the five representative vocations,

Freshman Choice and Senior Choice. Separate distributions both for the group showing a change and for the one showing no change in anticipated occupation were made on Freshman Choice and Senior Choice.

The means taken from the distributions of the difference scores (M_d) were examined to determine whether they indicated a trend, *i.e.*, a significant change from zero, or point of no change in interest. The ratio of the mean change to its standard error was the index used here. Upon the assumption that the general universe of which our sample is representative has a mean change in interest equal to zero, this method may be used to investigate the probability of obtaining by random sampling from such a general universe, a sample which would exhibit a mean change as great or greater than that obtained. This is made possible by accepting as an approximation, the standard deviation of the sample as equal to that of the general universe.

To determine the gross instability in interest disregarding directional trends, the quadratic mean of the difference scores was calculated for the five representative vocations, Freshman Choice, and Senior Choice. In considering the last two, the quadratic mean was computed separately for the group that showed a change in choice and the group that showed no change. The ratio between the quadratic mean change and the standard deviation of the freshman scores on the same vocation was used to facilitate comparisons between changes in interests occurring within an individual over the college period and the differences between the interests of individuals at a given time.

As an index to the consistency of vocational interest scores, the freshman-senior correlation coefficients were determined for the groups just listed. This coefficient is indicative of how effectively a man's interest as a senior may be predicted in his freshman year. The probabilities of obtaining a senior score of each letter grade or better with a given freshman score were determined, using the linear regression equations. To complete the analysis of the letter grades, the percentage of individuals with a freshman score of "X" who received a senior score of "Y" was tabulated for the vocations collectively and separately.

The relationship between the initial score and the amount of change is given by the freshman-difference correlations which were determined for the same groups. A measure of this relationship is desirable since the degree to which a man's interest pattern approximated that of the standardization group for a particular vocation when he enters college

as a freshman may be a significant factor in determining the nature of his change in interest during his college career. In addition to the correlation coefficients, eta, the correlation ratio was computed for the five vocations.

The scatter diagrams were examined by visual inspection for evident non-linear relationships. In the case of freshman-difference scores on Engineer, where the scatter diagram showed a parabolic tendency, a parabola of the form $d = AF^2 + BF + C$ was fitted to the data by the method of least squares.

To test the hypotheses that an increase in one vocational interest score may be the result of a corresponding decrease of interest score in another vocation or that instability is a function of the individual, intercorrelations between the difference scores for each vocation were obtained.

Further investigation of the factors influencing change was based upon a comparison of the highest and lowest quartiles of the difference score for each of the five vocations. For the cases exhibiting extreme changes the following data were recorded: Freshman and senior age, college, vocational choice with date of decision, extra-curricular activities, and professional fraternities.

INTERPRETATION OF RESULTS

A summary of the most important results is presented in Table III and the text that follows.

Examining the particular characteristics of the distribution of freshman and senior scores on the five vocations and the Freshman and Senior Choices it was found that Certified Public Accountant had by far the lowest mean value. The group studied approximated the interests of Lawyers more closely than any other occupational group with the exception of Office Clerk. The close relationship to the interests of Lawyers is probably an effect peculiar to the sample, since there is a greater probability of prospective law students remaining on the campus for three years or more. There is no apparent reason why the individuals comprising this sample should exhibit patterns of interest more nearly like those of office workers than those of any other representative vocation. Furthermore, the low weightings on the Office Clerk key may permit men in general to score nearly as high as the professional group which would also make for high scores here.

The analysis performed on the vocational choices showed that the group in each case was more interested in a chosen profession than in

some vocation arbitrarily selected for scoring. High freshman means were found in both the group with a change in choice and the group without a change between the freshman and senior years. The latter group had the highest freshman and senior mean values. In the case of senior scores on Freshman Choice in the group of twenty-eight cases which exhibited a change in choice between the two administrations of the tests, the loss in interest may be due to the rejection of one vocation in favor of another.

TABLE III.—SUMMARY TABLE

	Certified Public Accountant		Engineer		Lawyer		Teacher		Office Clerk	
	M	σ	*M*	σ	*M*	σ	*M*	σ	*M*	σ
Freshman	−5.83	1.91	−3.97	2.15	−3.29	1.11	−3.77	1.58	−2.65	1.87
Difference	.49	1.78	−0.68	1.20	.16	1.11	−0.04	1.49	−0.03	1.61
Senior	−5.34	2.09	−4.65	2.27	−3.13	1.21	−3.81	1.68	−2.68	1.78
RMS_d	1.84		1.37		1.12		1.49		1.61	
RMS_d/σ_f	.97		.64		1.10		.94		.87	
r_{fs}	.60		.85		.50		.58		.59	
r_{fd}	−.35		−.16		−.43		−.40		−.49	

	Freshman choice 66 cases		Change in choice 28 cases				No change in choice 35 cases	
			Freshman Choice		Senior Choice			
	M	σ	*M*	σ	*M*	σ	*M*	σ
Freshman	−2.48	1.84	−2.63	1.90	−2.79	2.00	−2.32	1.70
Difference	−0.05	1.31	−0.66	1.11	.36	1.60	.46	1.13
Senior	−2.53	2.06	−3.29	2.30	−2.43	1.80	−1.86	1.76
RMS_d	1.31		1.34		1.62		1.16	
RMS_d/σ_f	.71		.71		.82		.68	
r_{fs}	.78		.89		.36		.79	
r_{fd}	−0.18		.03		−0.54		−0.27	

The group which exhibits no change in vocational choice has mean values for both freshman and senior scores which fall within Strong's *B*

and *B plus* ranges. These are greater than any other means considered in the study. As interest in these chosen vocations becomes fixed during the college period as a result of either education, out-of-college information, maturity or any other cause, the subjects show an increase in interest test score from the freshman to the senior year.

Since the quartile deviation of each of Strong's standardization groups has been taken as the unit of measurement, the standard deviation of those groups will be approximately 1.5 in terms of *Q*-units, if the standardization distributions are approximately normal. When we compare the standard deviations of the freshmen with those of the standardization groups we find that the college freshmen are considerably more homogenous with respect to interests measured by the Lawyer scale than are Strong's group of lawyers. On Teacher, the homogeneity of interest is about the same, while freshmen are considerably more heterogenous than the standardization groups with respect to Certified Public Accountant, Office Clerk, and Engineer.

Of the five vocations, Engineer is the only one showing a significant trend and this is negative, an indication that there was a tendency for the group as a whole to become less interested in that profession. This characteristic may be due to the peculiarity of the sample. For example, of the entire group only seven gave Engineering as a vocational choice on the freshman blank and three of these had changed their choice by the senior year. If those who changed did so because of waning interests, their scores would be influential in the decrease. Of the nineteen cases showing the greatest increase in Engineer score, five individuals were in the Engineering College when seniors as compared to none of the nineteen cases showing the greatest decrease in interest, a factor which indicates that being in a certain college may facilitate the increase of interest scores related to that field. Trends in the other vocations, although they may be due to sampling, were negative in the case of Teacher and Office Clerk and positive in the case of Lawyer and Certified Public Accountant.

The only other significant trend (also negative) was found in the case of Freshman Choice in the group evidencing a change in anticipated occupation. There was no significant trend in the case of Senior Choice of this group, although the senior mean is higher than the freshman mean. This latter fact is also true for the group exhibiting no change. Whether the change in choice causes the decreased interest in the group showing the trend, or a decreased interest makes the man choose another profession can not be answered here. Analysis allows

the statement, however, that a change in vocational choice is paralleled by a decrease in interest score and the change in choice is accompanied by the greatest decrease found in the present investigation.

In every vocation, including Freshman Choice and Senior Choice, there appears a considerable degree of instability. Instability is dependent upon either a trend in interest test score, the dispersion of the changes in interests, or a combination of both. Table IV indicates that changes occurring within an individual between the freshman and senior years ranged from six-tenths as great, to changes equally as great as the differences between freshmen. A ratio between the quadratic mean and the standard deviation of the freshman scores permits the investigation of how the mean total change in interest scores which occurs within an individual during the college period compares with the differences between individuals at a given time.

TABLE IV.—QUADRATIC MEAN OF THE DIFFERENCES

	Certified Public Accountant	Engineer	Lawyer	Teacher	Office Clerk
RMS_d	1.84	1.37	1.12	1.49	1.63
RMS_d/σ_f	.97	.64	1.01	.94	.87

When the quadratic mean is thus referred to the freshman standard deviation on the respective vocations, Lawyer shows the greatest degree of instability with value of 1.01. This value represents an actual change without regard to direction which is a little greater than the standard deviation of the distribution of freshman scores on Lawyer. The significance of this score is seen if we consider that if an individual scores at the mean of the freshman distribution exhibiting the average gross change without regard to direction, his percentile rank in the freshman distribution might shift from .50 to .85, or to .15.

Engineer showed the greatest stability compared with the freshman dispersion. Lawyer was least stable when compared with freshman dispersion and highly stable when compared with the dispersion of Strong's standardization group. It is evident that the numerical values of the instability of the individual will appear large or small depending upon the homogeneity of the group with which they are compared.

So far as the author has been able to determine, no previous studies dealing with the Strong Vocational Interest Blank have been directed

toward an investigation of either trend or stability of score, but in each instance they have used as an index of permanence, the correlation coefficient showing the relationship between the first test and retest scores. This is a measure of consistency and may be high in the presence of large shifts in interest score. The correlation method has been applied in the present study, but its function has been shown to relate only to the study of the third attribute of change in interest score, that of consistency. The test-retest correlations of this study are compared in a table below with those obtained by Burnham[1] and Strong.[9]

TABLE V.—TEST-RETEST CORRELATIONS

	Florida three-year group	Yale three-year group	Stanford five-year group
Certified Public Accountant	.60	.66	.59
Engineer	.85	.78	.84
Lawyer	.50	.69	.77
Teacher	.58	.69	.66
Office Clerk	.59	...	.74

The freshman-senior correlation coefficient in each case denotes a degree of consistency unaccounted for by chance and is well within the range of the consistency of other tests frequently used on college groups in guidance work. Of the five representative vocations, Engineer shows the greatest degree of consistency with a coefficient of .85. Besides the tabled results, the freshman-senior correlation of .89 on Freshman Choice for the twenty-eight cases showing a change in choice is higher than any other coefficient obtained, so besides being the group of individuals showing the most marked trend, they are, so far as Freshman Choice is concerned, the most consistent. This together with the relatively high coefficient of .66 obtained from the group of seventy-one cases who expressed a senior choice, would tend to indicate that a definite choice may have an influence on consistency.

A correlation coefficient is based upon the assumption that a linear relationship exists between the two concomitant variables in question. Examination of the scatter diagram used in obtaining the freshman-senior correlation on Engineer revealed that a straight line here is not the best fit to the data. That the data approximated more nearly a parabolic function was shown by fitting a second degree parabola to

the relation between freshman score and difference scores. The index of correlation was .32. If the freshman-difference relationship is non-linear, then it can be shown that the freshman-senior relationship is non-linear. The freshman cases falling in the *A* and *C* ranges tend

TABLE VI.—PROBABILITIES OF A GRADE OF *X* OR ABOVE AS SENIOR WITH A FRESHMAN SCORE OF *Y*, COMPUTED FROM THE REGRESSION EQUATIONS

Certified Public Accountant: $\bar{S} = .66F - 1.50$, SE = 1.20

Freshman score	*A*	*B*+	*B*	*B*−
A	.17	.45	.76	.87
B+	.06	.25	.56	.72
B	.02	.11	.34	.51
B−	.01	.07	.25	.40

Engineer: $\bar{S} = .90F - 1.08$, SE = 1.20

Freshman score	*A*	*B*+	*B*	*B*−
A	.21	.51	.80	.98
B+	.06	.23	.54	.70
B	.01	.07	.26	.41
B−	.004	.03	.15	.27

Lawyer: $\bar{S} = .55F - 1.32$, SE = 1.04

Freshman score	*A*	*B*+	*B*	*B*−
A	.20	.45	.86	.94
B+	.09	.34	.71	.87
B	.03	.18	.51	.69
B−	.02	.12	.41	.59

Teacher: $\bar{S} = .62F - 1.47$, SE = 1.36

Freshman score	*A*	*B*+	*B*	*B*−
A	.21	.47	.75	.85
B+	.10	.30	.59	.72
B	.04	.16	.40	.55
B−	.03	.11	.32	.46

Office Clerk: $\bar{S} = .57F - 1.13$, SE = 1.44

Freshman score	*A*	*B*+	*B*	*B*−
A	.31	.58	.72	.89
B+	.19	.43	.69	.80
B	.10	.28	.54	.68
B−	.07	.22	.46	.60

Freshman Choice: $\bar{S} = .87F - .38$, SE = 1.30

Freshman score	*A*	*B*+	*B*	*B*−
A	.42	.72	.91	.96
B+	.20	.50	.75	.86
B	.06	.22	.50	.65
B−	.03	.14	.37	.52

to increase in score while those in the *B* range show instead a decrease. If the other investigators of this problem have found that a marked variance from an approximate straight line relationship exists in any of the occupational scales it has not been reported.

Reference is made to the consistency of letter grades since all investigations with which the results will be compared have accepted Strong's gross classification of scores. From the regression equation of senior score on freshman score and the standard error of estimate, the

TABLE VII.—PERCENTAGE OF STUDENTS MAKING A FRESHMAN GRADE OF *X* THAT MADE A SENIOR GRADE OF *Y*

Freshman score	Senior score	CPA		Engineer		Lawyer		Teacher		Office Clerk	
		Per cent	*f*	Per cent	*f*	Per cent	*f*	Per cent	*f*	Per cent	*f*
A	*A*	....	0	60.0	10	00.0	2	33.3	3	46.7	15
A	*B*	....	0	40.0	10	100.0	2	66.7	3	46.7	15
A	*C*	....	0	00.0	10	00.0	2	00.0	3	6.7	15
B	*A*	00.0	14	00.0	19	4.3	47	12.5	32	22.9	35
B	*B*	64.3	14	52.7	19	68.0	47	43.8	32	48.6	35
B	*C*	35.7	14	47.4	19	27.7	47	43.8	32	28.6	35
C	*A*	3.2	62	00.0	47	00.0	27	00.0	41	3.8	26
C	*B*	16.1	62	2.1	47	22.2	27	19.5	41	46.2	26
C	*C*	80.7	62	97.9	47	77.8	27	80.5	41	50.0	26

probability of obtaining a score equal to or better than each of Strong's letter grades when the freshman score was at the lower limit of the *A*, *B plus*, *B*, and *B minus* levels, respectively, was determined. Tabled results for each vocation are given in Table VI. Each row represents the lower limits of Strong's letter grades as an obtained freshman score, the columns give the lower limits of these grades as a senior, the entries indicate for an individual with a given freshman score the probability of his obtaining a score equal to or better than each of the letter grades as a senior. In addition, the reduced formula for the predicted value and the standard error of estimate is given in each case. These tables are based upon the best fitted straight line for the data. For the tables showing the empirical probabilities determined from the raw frequencies see Table VII.

The suggestion made concerning the possible existence of a non-linear relationship between freshman and senior score becomes valuable here. Table VI for Engineer gives a probability of .20 that an individual obtaining an *A* as a freshman will retain that rating or better his senior year. Examination of the empirical data, on the other hand,

shows that the individual has a probability of .60 of obtaining such a rating. Thus it is evident that there is a marked discrepancy between the prediction made from the actual data and that based upon the assumption of a linear function when such is not the best representation of the data.

In comparing the two tables giving the probability that a freshman making a given letter grade will make any particular grade as a senior, several points should be noted. Table VI gives the cumulative proportions making a particular senior grade or better while Table VII gives the proportions falling within a given letter-grade range. The values

TABLE VIII.—PER CENT OF THOSE MAKING A FRESHMAN SCORE OF *X* WHO MADE A SENIOR SCORE OF *Y*. ALL VOCATIONS. THREE INVESTIGATIONS

Percentage distribution of Freshman scores				Senior testing								
				A			*B*			*C*		
	S	*B*	*V*	*S*	*B*	*V*	*S*	*B*	*V*	*S*	*B*	*V*
A	4	..	..	63	..	..	36	..	..	1		
		5	..	..	35	..	..	57	..	..	8	
			8	..	..	47	..	..	50	..	..	3
B	29	..	..	14	..	..	67	..	..	19		
		33	..	..	9	..	..	62	..	..	29	
			39	..	..	10	..	..	56	..	..	35
C	67	..	..	1	..	..	21	..	..	76		
		62	..	..	1	..	..	20	..	..	79	
			53	..	..	2	..	..	18	..	..	80

in Table VI are dependent on the assumption of linear regression, while Table VII makes no assumption as to the regression. On the other hand, each value in Table VI has been determined by utilization of all the data, whereas, certain of the proportions in Table VII are determined on two to four cases. Both methods showed the letter grade of *C* to exhibit by far the greatest consistency, *B* is a little less consistent, while freshman score of *A* is least consistent of all. These results are consistent with the findings of Burnham and, to a lesser extent, with those of Strong. The greater consistency of *C* scores is due to the much greater range of interest represented by this single category.

Now if Table VII is reduced to the form where the senior score frequencies of the seventy-six University of Florida students on the five

vocations are expressed as percentages of the freshman score frequencies, the data may be compared directly with similar recapitulations made by Burnham[1] on his own data, and Strong's data. Strong's data, Burnham's data, and the data from the present study are represented in Table VIII by *S*, *B*, and *V*, respectively.

In Table VIII the columns of the freshman testing group add to one hundred, thus each entry here denotes the percentage of the total group of individuals who made that particular letter grade on the freshman test. The horizontal rows of the senior testing also add to one hundred for any one of the three studies. Thus of the five per cent of Burnham's group who made *A's* as freshman, thirty-five per cent made *A's*, fifty-seven per cent made *B's*, and eight per cent made *C's* as seniors.

When the letter grades on each of the five representative scales here are combined and compared with Strong's Stanford five-year graduate group together with Burnham's three-year Yale undergraduate group, the Florida scores are found to be much more comparable with the latter. This is evidence that the two undergraduate college groups are drawn from a similar universe.

Burnham pointed out that his data is highly comparable to Strong's with one exception. The discrepancy was in the Stanford finding that sixty-three per cent of the original *A* scores were likely to be repeated with only thirty-six per cent moving down to a *B*, as compared to only thirty-five per cent of the Yale group retaining an *A* and fifty-seven per cent dropping to a *B*. He has postulated a possible explanation on the basis that the Strong group was occupationally employed after graduation while the Yale group were still students during the test-retest interval. For this reason the patern of interests of Strong's *A* group may have had a tendency to become more nearly like the occupational standardization group while those of the students did not. The *A* scores of the Florida group confirm Burnham's finding in that they show a tendency to decrease.

It will be recalled that besides investigating what changes occurred in vocational interest scores over the college period, it was the further purpose of this study to note, where possible, factors which might influence the changes found to exist. This second aspect of the problem will now be considered.

Correlations between the freshman scores and the amount of change were computed on each of the five vocations, and for Freshman Choice and Senior Choice. The coefficients found were used as an index of the

existing relationship between two variables. It is realized that this correlation between initial status and gain is subject to the cumulative error from the errors of measurement both in the original and retest scores and some correction such as proposed by Thomson[10,11] or Cureton[3] should be applied. Since, however, the reliability of neither the test or retest was studied, the required information is not available for the correction and we shall use the raw correlations here as an approximation.

The coefficients showing the correlation between the freshman scores and the change in score are given in Table IX and compared with the coefficients showing the relationship between gains in occupational score and freshman scores found by Burnham in his three-year Yale group.

TABLE IX.—CORRELATIONS BETWEEN CHANGES IN SCORE AND FRESHMAN SCORES OVER THREE YEARS IN COLLEGE

	Florida data	Yale data
Engineer	−.16	−.34
Lawyer	−.43	−.36
CPA	−.35	−.35
Teacher	−.40	−.32
Office Clerk	−.59	

The two variables show low negative correlation, indicating that it is the lower scores which are likely to evidence the greatest increase. Practically every *A* score showed a decrease rather than an increase. The results for Engineer are atypical.

Inspection of the scatter diagrams used in obtaining the freshman-difference correlations revealed a very obviously non-linear relationship between the two variables on Engineer. Fifty per cent of the freshman *A's* increased, nearly as great a percentage of the *C*'s showed an increase, while seventeen of the nineteen freshman *B* scores showed a decrease. This feature accounts for the numerically low correlation coefficient which is based upon the assumption of a linear relationship, and predictions made on the basis of such a coefficient would be highly inefficient. So apparent were tendencies for the data to approximate a parabolic function in this case that the best fitting curve was determined by the method of least squares. The relationship between change in score and freshman score in the case of Engineer is then expressed by: $d = .030F^2 - .129F - .344$. The index of correlation

was .315, which is not significantly different from the value of η_{df} of .56 by Blakeman's short test for linearity ([6], p. 83). No greater values were found in the other four vocations, although the value for Certified Public Accountant was relatively high (3.923) and inspection indicated that non-linearity would probably have been revealed had any of the subjects attained *A* scores. In the vocational choice groups, the cases were too few to permit the use of a test for linearity. This last analysis is not only an index to the nature of the changes in Engineer scores, but it serves also to remind the investigators of this and similar problems that non-linear functions may occur which make the use of the correlation coefficient ineffective.

The possibility exists that the factors influencing a change in one interest score consistently affect changes in the other vocations. An alternative hypothesis is the existence of a trait of instability of interests. If the first is true, we may expect an increase in one vocation to be accompanied by a slight decrease in each other vocation, or a large decrease in some particular one. It might be, of course, that increases in one vocation are accompanied by increases in others. These would be shown by the intercorrelations of the difference scores. If the second hypothesis be true, then we would expect the individual showing large changes in one to show large changes (without regard to sign) in the others. This would be revealed by the intercorrelations of differences, ignoring sign.

TABLE X.—INTERCORRELATIONS OF DIFFERENCE SCORES

	Certified Public Accountant	Engineer	Lawyer	Teacher	Office Clerk
Certified Public Accountant...					
Engineer.....	.037				
Lawyer.....	.152	.153			
Teacher.....	−.087	−.196	−.324		
Office Clerk.....	.141	−.312	−.665	.312	

Table X gives evidence in specific instances in support of the first hypothesis—that changes in one vocation are at the expense of changes in other vocations. The most significant relationship between changes in interest was found in the case of Lawyer and Office Clerk, where increases in the former are at the expense of interest in Office Clerk.

It is impossible to state any causal relationship due to the lack of sufficient information.

In general, the intercorrelations obtained without regard to the sign of the changes were lower than those obtained by the former method. The relationships between Lawyer and Teacher and Lawyer and Office Clerk here were the only two that approached statistical significance. Since all of the coefficients are positive, we may conclude that, of the vocations studied, changes in one may be slightly related to changes in others, but as the coefficients are very small, such a relationship may be a reflection of the correlation with respect to sign. This disproves the second hypothesis proposed—that stability is a function of the individual.

TABLE XI.—INTERCORRELATIONS OF DIFFERENCE SCORES IGNORING SIGN

	Certified Public Accountant	Engineer	Lawyer	Teacher	Office Clerk
Certified Public Accountant...					
Engineer.....	.166				
Lawyer.....	.081	.097			
Teacher.....	.165	.030	.351		
Office Clerk.....	.046	.288	.353	.165	

If the age of an individual has anything to do with whether he will show a positive or negative change in interest test score during the college period it, at least, does not show a consistent influence throughout the five vocations studied. In the examination of the upper and lower quartiles of the difference scores on each vocation studied, the average senior age was about that of the entire group, with the exception of Certified Public Accountant which proved to be one year younger than the average. The freshman age on this vocation scale was also about a year below the average in the case of the extreme positive changes. This evidence would seem to indicate that Certified Public Accountant interests mature rather early. The freshman age was about a year below average in the case of the extreme increases on Office Clerk, whereas, the same is true for the extreme negative changes in the case of Teacher. Thus it may follow that an individual seeking guidance as a freshman who is younger than the average freshman of this group has a better chance of increasing his interest

score in Certified Public Accountant and Office Clerk, but a better chance of decreasing in Teacher interests.

Examining the changes in vocational choice between the test and the retest, the greater number of changes in choice seems to be found in the group showing the extreme positive changes in score with the exception of Office Clerk and Senior Choice, wherein there is a definite negative correlation between change in interest test score and instability of vocational choice.

The data show that, in general, if the individual is enrolled in the college which is most nearly related to the particular vocation in question, interest scores on that scale have a much better chance of increasing than of decreasing.

Changes in college between freshman and senior year are positively related to increase in Certified Public Accountant, Office Clerk, and Senior Choice score, and negatively related in Engineer and Lawyer. There were equal numbers of changes from one college to another for both the group showing the greatest increase and the group with the greatest decrease in score on Teacher scale.

Membership in professional fraternities related to the occupations studied is correlated with increased interests.

SUMMARY AND CONCLUSIONS

In an attempt to determine what changes occur in vocational interests during the college period and to note, where possible, factors which may influence changes, seventy-six University of Florida students who had taken the Strong Vocational Interest Blank as freshmen were retested with the same blank three years or more later. Both the test and retest blanks were scored on five representative vocations—Certified Public Accountant, Engineer, Lawyer, Teacher, and Office Clerk, together with the Freshman Vocational Choice and Senior Vocational Choice. Students changing their occupational preference between the two test administrations were treated separately in an analysis of changes in score on vocational choice. The raw scores were converted to quartile scores to permit comparison between vocations and distributions were tabulated for the senior, freshman and difference ($S - F$) scores.

The investigation of the permanence of vocational interests between the freshman and senior years of college included the consideration of three distinct attributes of change in vocational interest test scores.

(1) *Trends* in interests were determined by investigating the mean of the differences in interest score between the two tests. A significant negative trend was found in Engineer scores. Considering the group of individuals evidencing a change in anticipated occupation, no appreciable change in score over the three years appeared in the case of the Senior Choice, but the Freshman Choice exhibited a consistent decrease in score which proved to be statistically the most significant trend found in the entire study. Of the group of individuals evidencing no change in anticipated life's work there is a significant increase in score over the college period on the vocations of their choice.

(2) *Total instability* of interests was measured by the quadratic mean of the differences. Considerable instability was found in all of the vocations studied. Changes within an individual ranged from six-tenths as great to equally as great as the differences between freshmen. Engineer, in spite of the large mean change, shows the greatest stability when the quadratic mean of the differences is compared with the freshman dispersion.

(3) *Consistency* was studied directly from the freshman-senior correlations. Group predictions may be successfully based upon the relationships found to exist between freshman and senior scores. Results indicate that the experiences effecting interests acquired during the college interval are largely determined by experiences acquired previous to college entrance. Chosen vocations show a higher degree of consistency than other vocations and the probabilities of attaining any letter grade as a senior with a given freshman grade, as determined by the regression equation, were higher for Freshman Choice than for any of the representative vocations.

Neither trend nor stability have been reported or suggested in previous studies on permanence of interests and the consistency of scores is the only attribute of change which is common to other investigations.

There was evidence in the freshman-senior correlations of non-linear relationships in the case of Engineer and Certified Public Accountant. Significant non-linearity has not been mentioned by the other investigators of permanence of interest. A parabolic regression of senior Engineer score on freshman score gave a correlation index of .32.

The probabilities that an individual with a given freshman grade will, as a senior, reach or exceed any letter grade, based upon the assumption of a linear regression between freshman and senior scores,

have been presented for the various vocations. The percentage making a given freshman score who actually made each score as a senior has been given in a separate table. There is a tendency for high initial scores to show a decrease.

Investigation of the factors which may influence changes in vocational interests revealed that, in general, changes in score are negatively correlated with freshman scores on the various interest scales. The two variables in Engineer were related by a second order parabolic function. This analysis serves to emphasize the necessity of testing for the existence of curvilinear relations.

Changes in interest on one vocational scale are related to changes in other specific scales. The most striking example is found in the case of Lawyer and Office Clerk where increases in score on one scale are paralleled by decreases in the other. There is no evidence that stability is a function of the individual.

The highest and lowest quartiles of the difference scores were compared in regard to external criteria which might influence change in interests. The effect of age is not consistent throughout the different vocations. Changes in interest test scores between the administration of the two tests are in general positively related with instability of vocational choice. The individual has a further chance of increasing in interest test score if he is enrolled in the college which is most nearly related to his senior vocational choice. Changes in college between freshman and senior year are related to increases in score on Certified Public Accountant, Office Clerk, and Senior Choice, whereas, it is correlated with decreases in Engineer and Lawyer. Membership in professional fraternities is related to increases in interests.

BIBLIOGRAPHY

1. Burnham, P. S.: *Stability of Interest Test Scores.* Ph. D. thesis, Yale University Library, 1935.
2. Bureau of Vocational Guidance and Mental Hygiene. "Bulletin." *University Record, University of Florida,* Vol. XXIX, 1934, pp. 167–180.
3. Cureton, E. E.: "A modification of Thomson's Formula for the Correlation Between Initial Status and Gain and Its Standard Error." *J. Exp. Psychol.,* Vol. XIII, 1930, pp. 358–364.
4. Finch, F. H.: "Permanence of Vocational Interest." *Psychol. Bull.,* Vol. XXXII, 1935, pp. 682.
5. Fryer, D.: *The Measurement of Interests.* New York: Henry Holt and Co., 1931.
6. Holzinger, K. J.: *Statistical Methods for Students in Education.* Boston: Ginn and Company, 1928.

7. Rock, R. T., Jr.: "A study of the Constancy of Responses to the Items of the Strong Vocational Interest Blank." *Psychol. Bull.*, Vol. XXXI, 1934, pp. 705–706.
8. Strong, E. K., Jr.: *Manual For Vocational Interest Blank for Men.* Stanford University: Stanford University Press, 1937.
9. ———: "Permanence of Vocational Interests." *J. Educ. Psychol.*, Vol. XXIV, 1934, pp. 336–344.
10. Thomson, G. H.: "A Formula to Correct for the Effects of Errors of Measurement on the Correlation of Initial Values With Gains." *J. Exp. Psychol.*, Vol. VII, 1924, pp. 321–324.
11. ———: "An alternative Formula for the True Correlation of Initial Values With Gains." *J. Exp. Psychol.*, Vol. VIII, 1925, pp. 323–324.

ITEM VALIDITY AS A FACTOR IN TEST VALIDITY*

ABRAHAM KROLL

Benjamin Franklin High School, New York City

There has been uncritical acceptance of the apparently logical contention that valid items when combined into a test form a valid test. By extension of this idea it has been further assumed that the most valid items would combine to form the most valid test. The publication in 1934 of Smith's[6] experimental scrutiny of this assumption was the first in this field. Little work relating to this topic has been reported in the literature, only one major investigation[5] directly concerned with this problem having appeared since Smith's study. Apparent inadequacies in both these studies provoked the present undertaking which concerns itself with answering the question, "Do the most valid items form the most valid test?"

It was felt that this question had not been answered adequately by the previous studies because the numbers of cases on which the test-item validities had been computed were insufficient to assure stability, because defective experimental design influenced the results to an unknown but perhaps significant degree, and because of the inadequacy of the criteria employed. A closer approximation to the true answer to this question was sought in the present investigation by following a superior plan of procedure which avoided the principal shortcomings of the two earlier studies.

PREVIOUS STUDIES

Smith[6] investigated the problem—Is the labor involved in calculating item validity coefficients repaid through the improvement of results?—*i.e.*, "Will the more valid items furnish a definitely superior test?" He sets out the facts in regard to the comparative validities of his subtests in his Table V. The principal criticisms which may be levelled at Smith's study are, (1) Smith's population was highly selected and the material for his test items was drawn from the well worked over and highly validated items in the CAVD pool at Teachers College; (2) he computed the biserial item validity coefficients from the data obtained from the three hundred seventy pupils in 1930, assuming that these coefficients could then be applied to the same items for

* This paper is a condensation of a dissertation accepted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at Fordham University.

subsequent groups, though the test items were rearranged and the numbers increased. The lack of consonance of the results may be due to this assumption; and (3) his criterion test was composed of the one hundred twenty items remaining after selecting the eighty items used in his subtests (see Fig. 1, p. 431, for graphic representation of Smith's method of selection of criterion), and that this method of selecting the

TABLE I.—CORRELATION OF EACH SUBTEST WITH THE CRITERION: 1930, 1931 AND 1932
(Smith's Table 5)

Subtest	Correlation of sub-test with the criterion			
	1930 data, $N = 370$	1931 data, $N = 542$	1932 data, $N = 527$	Weighted average for all three years, $N = 1433$
"Best" subtest	.8471	.8194	.9202	.8717
"High" subtest	.8691	.8636	.8972	.8782
"Low" subtest	.8009	.8140	.8561	.8276
"Worst" subtest	.4304	.4119	.4951	.4478

criterion test placed his "best" subtest at a disadvantage from the viewpoint of correlation technique. By noting the arrangement of the "high" and "best" subtests with relation to the "criterion" test, it can be seen that the "high" subtest has more in common with the criterion test than the "best" subtest. This is so because the "high" subtest is extracted from the very body of the criterion, whereas the "best" subtest is outside the criterion. One could thus anticipate that the "high" subtest would be favored in relative magnitude of the correlation coefficient, and the "best" subtest be handicapped. It is significant to note that Table I presents some evidence of the truth of this surmise.

Long and Sandiford extended the work begun by Henry[2] in an attempt to throw light upon the applicability and worth-whileness of item-validity techniques toward constructing tests of greater validity. They admit that there is spuriousness in their correlation coefficients and that their population of one hundred forty-four was highly selected and homogeneous. They set up their fifty-item tests and correlated the scores against the criterion of total scores on four hundred eighty-two items which included the fifty items in the subtest. When they

used biserial-*r*, or the Kelley Method[3] for selecting the items, they found the most valid test was not the one which contained the most valid items. However, when they selected items by the Upper-Lower Thirds Method or by the McCall-Long-Bliss Method, the best test was composed of the best items. They ascribed this variation to "accidents of circumstances rather than to the property of the techniques themselves."

Another factor which Long and Sandiford did not discuss in their presentation was the difference between the method which they employed in spacing their subtests and the method which Smith used. Smith's subtests are spaced at intervals with numbers of highly valid items omitted, so that the differences are emphasized. The Long and Sandiford items, when grouped into subtests, are taken without intervals. Their fifty-item subtests are adjacent groups of fifty items. If they do not find a sharp difference between adjacent subtests, their results do not necessarily corroborate Smith's, and their findings should not be cited as supporting Smith's conclusions.

Thus it can be seen that both these studies used certain procedures that could be improved upon. The number of cases on which the items were to be validated could be increased. Smith used three hundred seventy, Long and Sandiford used one hundred forty-four. Both studies employed highly selected populations. Smith used very much worked over and carefully validated items from the CAVD pool. These conditions are not typical of those faced by the average test constructor who must prepare a test for a general population from untried, newly devised items. Both studies permitted the spuriousness of self-correlation to enter into the relative magnitudes of their coefficients. Each study presents results that are not consistent.

PROCEDURE

The data for the present investigation were obtained from objective tests in Spanish administered to all the students, boys and girls, registered in the sixth term of Spanish in all the public high schools of New York City.* A criterion test of knowledge of Spanish was given in three parts. The items for these three parts were obtained from the experimental, objective-type Regents Examination in Spanish for June 1925 developed by Ben D. Wood[7] of Columbia University. In

* Permission to administer the tests to the public-high-school students was obtained through the kind coöperation of Dr. Lawrence H. Wilkins, Director of Modern Languages for the Board of Education of New York City.

the published analysis of the examination, Wood reports the order of difficulty of each item within each part. As administered in this study, the items were rearranged in the order of reported difficulty, from the least difficult to the most difficult. In addition to the criterion test, eight objective, multiple-choice tests were administered, each consisting of fifty items, dealing with Spanish vocabulary. The four hundred vocabulary items were selected from three sources, the New York State Syllabus in Modern Languages,[8] A Graded Spanish Word Book[1] and A Basic List of Spanish Words and Idioms.[4] The words were so selected that none duplicated any that appeared in the criterion test.

The total number of cases available for study was eleven hundred fifty-four. From these, ninety were eliminated because of the extremely high average criterion scores caused by background knowledge of Spanish, Portuguese or Italian. Of the remaining ten hundred sixty-four in the population, three hundred ten cases were incomplete, *i.e.*, in each instance, one or more parts of the criterion or of the experimental vocabulary tests had been omitted. This left seven hundred fifty-four cases for whom there were complete data.

The distribution of the criterion test scores, consisting of the total scores derived from a vocabulary test, a comprehension test and a grammar test, (hereafter referred to as the VCG test), was checked for normality and found not to deviate significantly from normal. With the aid of punched cards and International tabulating equipment the biserial correlation coefficient of each of the four hundred experimental vocabulary test items was computed, using the VCG scores as the criterion. The distribution of these coefficients is shown in Table II. The range of the coefficients is from $-.62$ to 1.10, the mean being .236 with a standard deviation of .133.

The four hundred experimental vocabulary items were now arranged in descending order of biserial correlation coefficient, renumbered consecutively from 1 to 400, and arranged in six groups. All the even-numbered items were grouped to form the "new criterion test," in order to secure a sample of items spread evenly over the entire range of validity. Attention is here directed to the basic difference between the new criterion test as selected in this study and the criterion tests as selected by Smith or by Long and Sandiford. Smith sampled only part of the range of biserial coefficients for his new criterion test, while Long and Sandiford used the remaining items after "randomly" choosing their subtests. The present study systematically samples

the full range of biserial coefficients to obtain the items for the new criterion test.

The remaining two hundred odd-numbered items were divided into five groups to duplicate the Smith technique. That is, the odd-numbered items 1–39 were grouped to form the "best" subtest; those numbered 121–159 were grouped to form the "high" subtest; those numbered 241–279 were grouped to form the "low" subtest; those numbered 361–399 were grouped to form the "worst" subtest; the remaining odd-numbered items (41–119; 161–239; 201–359) were grouped to form a test which will be referred to as to "Smith Criterion Test." The selection of the items and the arrangement of the tests are indicated in Fig. 1.

TABLE II.—DISTRIBUTION OF BISERIAL COEFFICIENTS OF CORRELATION FOR FOUR HUNDRED ITEMS (N = 754)

Size of coefficient	f	
1.10– 1.19*	1	
1.00– 1.09*	1	
.90– .99		$\cdot M$ = .236
.80– .89		
.70– .79		$Md.$ = .240
.60– .69		
.50– .59	2	SD = .133
.40– .49	13	
.30– .39	90	
.20– .29	156	
.10– .19	99	
.00– .09	28	
−.10– −.01	6	
−.20– −.11	2	
−.30– −.21	1	
−.40– −.31		
−.50– −.41		
−.60– −.51		
−.70– −.61	1	
	400	

* The writer is indebted to Dr. M. W. Richardson, of the University of Chicago, for the following discussion of the cause for the value of biserial r exceeding unity:

"The biserial coefficient has an upper limit of unity only when the y-distribution (the continuous one) is normal. This point is missed in most treatments of the subject, presumably because the authors have been too anxious to make the point that the x-distribution (the dichotomized one) is assumed to be normal. If

THE DATA

The scores on the tests formed by "pooling" specific items provided the basic data upon which this investigation depends. The ranges, means and standard deviations of the biserial correlation coefficients are given in Table III. It should be noted that each of the four

TABLE III.—RANGES, MEANS, AND STANDARD DEVIATIONS OF THE BISERIAL COEFFICIENTS OF THE ITEMS FOR EACH OF THE "POOLED" TESTS

"Pooled tests"	Number of items	Range of biserial *r*'s	Mean biserial *r*	SD of the biserial *r*'s
"New criterion"	200	1.056– –.619	.235	.140
"Best" subtest	20	1.104– .364	.445	.095
"High" subtest	20	.292– .266	.279	.002
"Low" subtest	20	.221– .190	.205	.003
"Worst" subtest	20	.108– –.272	.003	.040
"Smith criterion"	120	.361– .108	.236	.060

"pooled" twenty-item tests has a range of biserial correlation coefficients not overlapping the ranges of the other three pools. Also the limited range of the "Smith Criterion" coefficients, .361 to .108, can be

---

the continuous or *y*-distribution is *flatter* than normal, and the relationship is sufficiently high, the computed coefficient may be greater than one. In such cases, of course, the biserial coefficient should not be used, if one considers the matter strictly. However in the application you suggest, I do not think any harm is done.

"Suppose that the *y*-distribution is concentrated equally in two points, separated by a distance of d. We then have $\frac{1}{2}N$ points with a *y*-value of $y' + d$. The mean is $y + \frac{1}{2}d$. The deviation values are $\frac{1}{2}d$ and $-\frac{1}{2}d$, respectively. The standard deviation is

$$\sqrt{\frac{\frac{N}{2}\left(\frac{1}{2}d\right)^2 + \frac{N}{2}\left(-\frac{1}{2}d\right)^2}{N}}, \text{ which reduces to } \sqrt{.25d^2} = 0.5d$$

The maximum biserial r is then

$$\frac{M_p - M_t}{\text{SD}} \cdot \frac{pq}{z} = \frac{d}{0.5d} \cdot \frac{.5 \times .5}{.3989} = 1.26$$

This maximum occurs at $p = .5$, and is independent of the distance between the two point-clusters. Any distribution which approaches the point-distribution will have a larger value of r than has a normal distribution."

compared with the full extent of the range of coefficients in the "New Criterion" used in this study, 1.056 to −.619. This is further emphasized by a comparison of the magnitudes of the standard deviations of

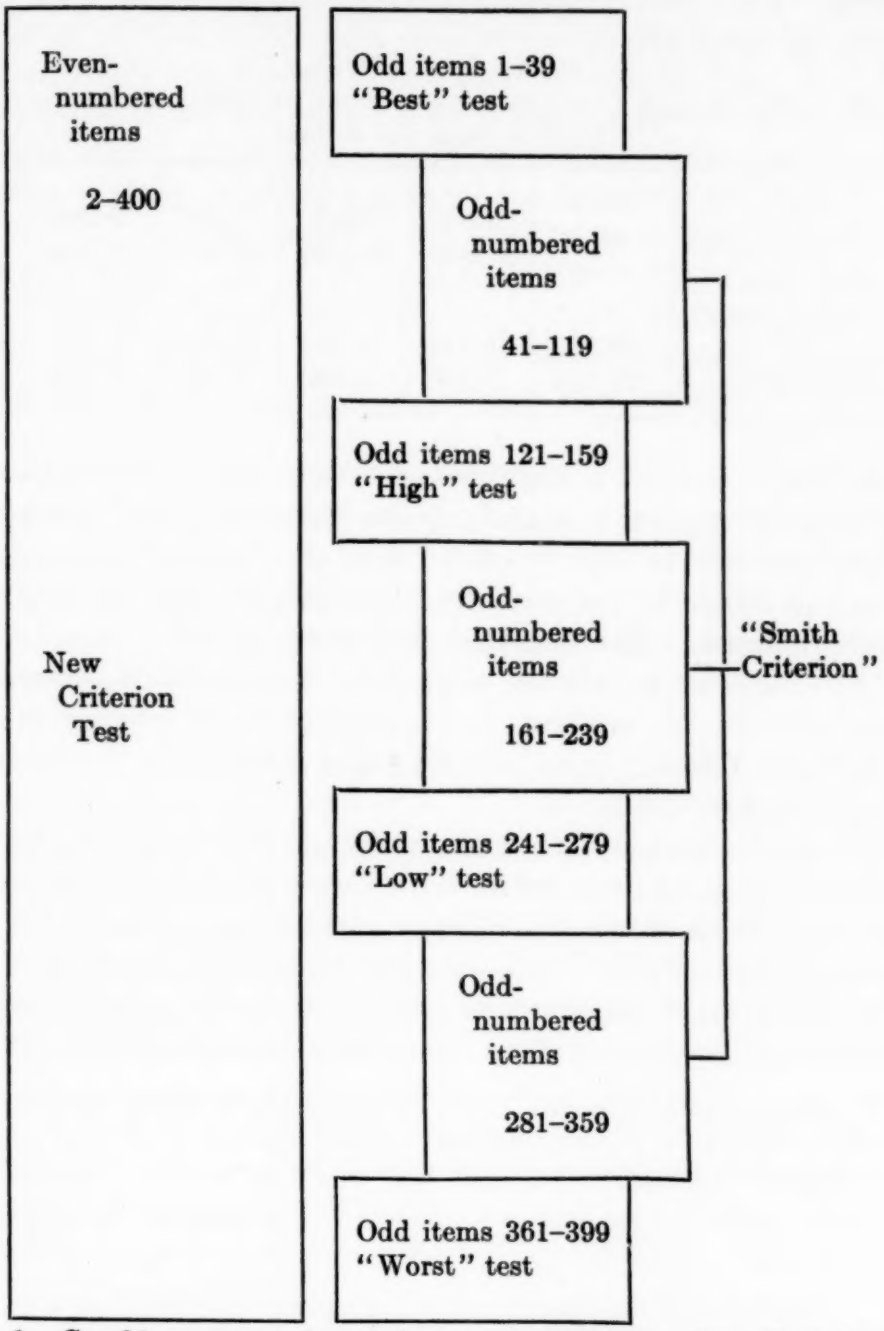

FIG. 1.—Graphic representation of the grouping of items to form the tests used.

the ranges of the biserial *r*'s of these two criteria. Despite the fact that the means of the two tests are alike, the standard deviations differ markedly.

The scores on each of the six pooled vocabulary tests for each of the seven hundred fifty-four students were ascertained by tabulating a specially prepared set of punched cards on International tabulating machines. The means and the standard deviations of the distributions are given in Table IV.

TABLE IV.—MEANS AND STANDARD DEVIATIONS OF THE POOLED VOCABULARY TEST SCORES (N = 754)

	"New criterion"	"Best" subtest	"High" subtest	"Low" subtest	"Worst" subtest	"Smith criterion"
Mean	125.703	14.831	15.414	13.076	11.901	74.721
Standard deviation	22.544	2.990	2.830	2.826	1.976	13.113

The "high" subtest is slightly easier than the "best" subtest, the "best" subtest in turn is slightly easier than the "low" subtest, and the "low" subtest is slightly easier than the "worst" subtest. The difference in difficulty between the most difficult and the easiest subtests is 3.5 items. The standard deviations of the "best," "high" and "low" subtests are almost alike, 2.99, 2.83 and 2.83, respectively. In these respects, the subtests in the present study employ data that are materially different from corresponding data in the Smith and the Long and Sandiford studies.

In order to ascertain the relative validities of these pooled tests, the scores for each were correlated with the scores on each of three separate criteria, the VCG scores, the New Criterion scores and the Smith Criterion scores. Table V presents the significant correlation coefficients. Using the VCG scores as the criterion, the relative validities of the subtests represented by the correlation coefficients (.729, .595,

TABLE V.—CORRELATIONS OF THE BEST, HIGH, LOW AND WORST SUBTESTS WITH THE CRITERIA (N = 754)

Criterion	*VCG* (1)	New (2)	Smith (3)
Best subtest (4)	.729	.763	.732
High subtest (5)	.595	.702	.712
Low subtest (6)	.468	.638	.648
Worst subtest (7)	.361	.309	.292

.468, .361) indicate that the "best" subtest has the greatest relative validity, and the "high," the "low" and the "worst" subtests have progressively lesser validity. When the new criterion is used, the correlation coefficients representing the relative validities of these subtests are all higher (.763, .702, .638 and .309, respectively), with the exception of the one representing the correlation between the "worst" subtest and the new criterion. This is slightly lower than it was when the VCG scores were used as the criterion. The "best" subtest retains its position of highest relative validity and the "high," "low" and "worst" subtests again are found to have progressively lower correlation coefficients in the order named. When the criterion is selected according to the Smith technique, the relative validities of the subtests retain the same order, but the magnitudes of the coefficients are different, (.732, .712, .648 and .292, respectively).

THE RESULTS

These results are uniformly consistent in so far as the three separate procedures with the four subtests indicate the validity coefficients to have the same relative order of magnitude, in each case the best items forming the best test and the better items forming the better test. In this important respect the present study differs from the results reported by previous investigators.

To determine whether the differences between the various coefficients were reliable, or whether they were due to sampling, the standard errors of the differences and the critical ratios were computed. These are presented in Table VI. Here we find seven of the nine critical ratios to be between 2.5 and 9.9, indicating that the obtained differences may be considered statistically significant. Another of these ratios is about 2.3 and one has a value of only 1.0. The general trend of the critical ratios in the direction of 2.5 or more may be interpreted as evidence that reliable differences have been presented.

Long and Sandiford interpreted Smith's results as a corroboration of their own. They ascribed the relative differences in the validity coefficients of the subtests in Smith's study and in their own, to the decrease in variability of the scores of the subtests. This decrease in variability, they said, was due to the progressively increased difficulty of these subtests. Where the former studies found the subtests to be progressively more difficult, and the scores, therefore, progressively less variable, the present study offers no indication of significant differences either in the difficulty of the subtests or in the variability of the scores.

TABLE VI.—DIFFERENCES BETWEEN CERTAIN CORRELATION COEFFICIENTS IN TABLE V, STANDARD ERRORS OF THE DIFFERENCES AND THE CRITICAL RATIOS
1—VCG Crit.; 2—New Crit.; 3—Smith Crit.; 4—Best Test; 5—High Test; 6—Low Test; 7—Worst Test

Correlation coefficients	r_{rr}	Difference (a)	SD difference (b)	Ratio a/b
$r_{14}-r_{15}$	.506	.134	.021	6.381
$r_{24}-r_{25}$	.290	.061	.021	2.905
$r_{34}-r_{35}$	.354	.020	.020	1.000
$r_{15}-r_{16}$	.398	.127	.029	4.379
$r_{25}-r_{26}$	.129	.064	.028	2.286
$r_{35}-r_{36}$	.294	.064	.026	2.462
$r_{16}-r_{17}$	.159	.107	.039	2.744
$r_{26}-r_{27}$	.145	.329	.037	8.892
$r_{36}-r_{37}$	.157	.356	.036	9.889

The "best" subtest is more difficult than the "high" subtest by .5 vocabulary item and less than two items more difficult than the "low" subtest. In turn, however, the "worst" subtest is only one item more difficult than the "low" subtest. The standard deviations of the "best," "high" and "low" subtest scores are between 3.0 and 2.8, while that for the "worst" subtest is 2.0, and the differences among the standard deviations are not statistically significant. Therefore, it cannot be said that the differences between the correlation coefficients which indicate the relative validity of these subtests are caused by variability of the scores. The causal factor which contributes to the differences in the validity coefficients of the tests must be the differences in the test-item validities of the elements that composed the subtests.

If the use of the biserial r technique resulted in the selection of better items in this study, why did it not do likewise in the Smith and in the Long and Sandiford studies? The inadequacy of the Smith criterion, the small number of cases in the Long and Sandiford study, the unknown effect of spuriousness which entered the correlation coefficients in both studies, the highly selected populations employed and the previously validated items used, may account for the difference. The Smith criterion was inadequate because it employed only part of the total range of biserial correlation coefficients; twenty per

cent of the range of coefficients was not included at all (ten per cent at each extreme). This tended to reduce the magnitude of the difference between the validity coefficients of the "best" subtest and the "high" subtest and also to reduce the reliability of this difference. In the Long and Sandiford study the total number of cases involved in the population was one hundred forty-four. The present study has avoided spuriousness, has sampled the full range of the biserial coefficients to form the new criterion, has used a large number (seven hundred fifty-four) of not specially selected cases, and has dealt with untried new test items.

To indicate further the relative improvement which test item validity techniques may afford for the selectivity of test items, the intercorrelations between the subtests are presented in Table VII. If one

TABLE VII.—INTERCORRELATIONS OF THE SUBTESTS (N = 754)

	"Best" subtest	"High" subtest	"Low" subtest
"High" subtest	.607		
"Low" subtest	.496	.521	
"Worst" subtest	.185	.243	.237

is to consider each twenty-item subtest as a secondary criterion for comparing the other twenty-item subtests, the intercorrelations can be used to indicate their relative validities. The "best" subtest correlates with the "high" subtest .607, with the "low" subtest .496 and with the "worst" subtest .185. Here we see the descending order of relative validity. If the "high" subtest is to be taken as the criterion, it correlates with the "low" subtest .521 and with the "worst" subtest .243. The "low" subtest is relatively better than the "worst" subtest. The intercorrelation between the "low" subtest and the "worst" subtest is .237 and may be taken as an indication of the low validity of the material in these two tests.

SUMMARY

Two important studies addressed to the question of the value of employing item validity techniques for the improvement of test validity, supported each other in their published statements and came to the seemingly paradoxical conclusion that little gain is to be anticipated by employing item validity procedures in order to improve test validity. Both studies appeared to lack adequate numbers of cases for validating

items reliably; both studies permitted spuriousness to affect their correlation coefficients to an unknown but possibly significant extent; both studies used highly selected populations; both studies used items which had been validated by others; and both studies used experimental designs which tended to reduce differences between their correlation coefficients. Because of these numerous defects, the present study was undertaken to determine more accurately the answer to the question, "Do the most valid items form the most valid test?"

To answer this question the author presented a study which employed seven hundred fifty-four subjects, thus providing a relatively large number of cases. At every point spuriousness was avoided. The VCG criterion test was administered for use as a separate criterion in the biserial r technique for computing item validity coefficients. The population was not specially selected but included all the students registered in all the public high schools in New York City who were studying the sixth semester of Spanish. The items in the experimental tests were new items constructed by the writer. The experimental design employed in determining the relative test validities gave no advantage to any subtest. The results derived have shown conclusively that the best items when grouped form the best test.

BIBLIOGRAPHY

(1) Buchanan, Milton A.: *A Graded Spanish Word Book, a Compilation*, Third Edition. Toronto, Canada: The University of Toronto Press, 1936.

(2) Henry, Lorne J.: "A Comparison of the Difficulty and Validity of Achievement Test Items." *Journal of Educational Psychology*, Vol. xxv, 1934, pp. 537–541.

(3) Kelley, Truman L.: "The Selection of Upper and Lower Groups for the Validation of Test Items." *Journal of Educational Psychology*, Vol. xxx, Jan. 1939, pp. 17–24.

(4) Kenniston, Hayward: *A Basic List of Spanish Words and Idioms.* Chicago, Ill.: The University of Chicago Press, May, 1933, pp. 1–68.

(5) Long, John A. and Sandiford, Peter, and Others: *The Validation of Test Items*, Bulletin #3, Department of Educational Research, Toronto, Canada: University of Toronto Press, 1935.

(6) Smith, Max: *The Relationship Between Item Validity and Test Validity*, Contributions to Education No. 621, New York: Bureau of Publications, Teachers College, Columbia University, 1934, pp. 40.

(7) Wood, Ben D.: *New York Experiments with New-Type Modern Language Tests.* New York: The Macmillan Company, 1928, pp. 230–249.

(8) ———: *Syllabus in Modern Languages*, (1933 reprint with minor revisions). Albany, N. Y.: The University of the State of New York Press, 1933, pp. 113–130.

THE PROBLEM OF LEARNING READINESS

J. A. LYNCH

The Rice Institute

One of the fundamental problems connected with the theory of curriculum is to determine as nearly as possible the order in which teaching objectives should succeed each other. The purpose of this paper is to state that problem as it appears from a perspective which I shall attempt to clarify by making some comparisons with Dewey's philosophy of curriculum.

The problem, as I conceive it, reduces down finally to that of giving an exact definition of a prerequisite. This does not imply that prerequisites are themselves necessarily definite and fixed. In the nature of the case, the statement of this problem cannot be final, since the principles upon which it is based are highly theoretical. But the more accurately the principles upon which it rests are stated, the more readily criticisms can be made and understood.

Therefore, I find it economical of space to begin with a brief statement of principles which I have attempted to justify in some other papers to which references will be made. These principles will be found to depart in certain respects from Dewey's conception of the learning process. To anticipate the argument, I wish to put a little more emphasis upon the quantum or wholeness character of both the learning process and the learning product than is suggested by the association type of psychology which qualifies the Dewey philosophy of education. The practical bearing of this change of emphasis will be apparent. The emphasis upon quanta which change by increasing clarity in both outline and detail rather than by merging and blending with each other applies also to aims. It seems to me that too much emphasis upon the fluid character of the learning process and the projected ends and aims which subtend and motivate it confuses certain educational issues and gives rise to insuperable difficulties in practice. It seems to be more consistent with reality as well as of practical utility to shift the emphasis from ends and aims so conceived to goals which can be represented as changing more in degrees of clarity than in fundamental pattern. For instance, the recurring contemplations of a plan to own a home should result in increased clarity rather than in transition to something fundamentally different, through association, as things do in *Alice in Wonderland*. A child who discovers a pattern by which he can make a boat will try it out repeatedly over a period of

time before he perfects it to the stage at which it no longer interests him; afterwards, he will occasionally make a boat or two by that pattern as if by way of review, often with some added parts.[1] When we look upon these experiences as recurrences of the same thing, life takes no more of a rhythmical character than it appears to have from the association point of view. It is often pointed out that a boy readily learns a principle, such as that of the electro-magnet, when it is discovered to have a close connection with some of his practical interests. There is an element of truth in this, because the practical interest, if mature, constitutes a prerequisite, as the latter is defined in this paper. But there is more to the problem than merely getting the idea into the consciousness of the learner. The young scientist, in whom the idea comes to life, will return to this principle repeatedly as the pattern of his creative play. He will model as many motors, etc., as the resources at his command make possible or until the idea matures. These resources may not be much more than pencil and paper with which to draw schematic pictures and books which contain pictures and diagrams. Succeeding levels of interest in the principle result whenever the boy gets control of new materials which make possible the expression of the principle in more complex situations. This sort of thing, referred to above as a learning quantum, is brought into prominence in the following discussion; it is applied to goals and to typical life situations as well as to ideas.[2]

The conception of a prerequisite for a particular learning product, as described below, depends upon the manner in which the distinction is drawn between the learning product being formed and the conditions for it. This distinction, as I conceive it, can be made clear by the use of a formula,

$$\sum_{i=n}^{i=\infty} a_i = A_\infty,$$

which is explained as follows:

---

[1] I have observed this in the behavior of my children.

[2] Whitehead, A. N.: *The Principles of Natural Knowledge.* Cambridge University Press, 1925. The last chapter, on "Rhythm," is a profound treatment of the relation of pattern to rhythm.

Bosanquet, B.: *The Principle of Individuality and Value.* The Gifford Lectures for 1931. Lecture II, "The Concrete Universal," states a point of view which is suggestive of some of the ideas developed below, in this paper. See especially page 37.

The term, a_i, is any member of a series of situations to which A_∞, a matured learning product, is relevant.[1] I maintain that after a sufficient number of practices, a learning product matures, and that it cannot significantly influence further learning until that stage is reached. This emphasis upon maturity is impossible from the Dewey point of view which represents the succeeding practice situations and their attending learning products as changing in such a manner as to leave nothing of a fundamental nature to be mastered through repetition.[2] As an illustration of the above formula, the law of inverse squares is relevant to an infinite number of situations in nature which are alike only with respect to the meaning given them by this law. The law selects out and defines a series of situations to which it is relevant. Some of these are capable of being used as practices for this principle as a consciously-recognized teaching objective. It is better to think of this law as the constant meaning in all the members of the series, although the insight into it may vary until maturity is reached. At that stage, the learning product extends its interpretative pattern to all other relevant situations. When maturity is reached, the memory, which might have been an important factor in connection with the formative practices, is no longer the key to knowing or not knowing.[3] The various possible verbal, mathematical, and schematic statements of this law also belong to the series expressed above as a summation. The essential difference between these and the physical situations which express the law consists in their being made up largely of objects used in their referential rather than in their existential character. Pure symbols, such as spoken words, are typical of the referential; but some objects composing a meaningful situation are used both existentially and referentially. Some are composed of objects of both extreme types. In general, the more a series has been influenced by human activity, especially by attempts at communication and expression in the past, the greater will be the tendency for the referential to predominate over the existential meanings. It is obvious also that learnings which depend largely upon introspection and upon

---

[1] See my article, "A Quantitative Description of the Stimulus." *Psychometrika*, July, 1938; also, "The Transfer of Training," *The Journal of General Psychology*, January, 1939.

[2] I have criticized this aspect of Dewey's philosophy in an article entitled, "A Criticism of Dewey's Conception of the Stimulus," accepted for publication by the *Philosophical Review*, under Discussion.

[3] The rôle of memory is defined more specifically later.

insights into the less tangible human relationships for their data will be embodied in symbols. The latter type of learning situations is predominantly humanistic, *i.e.*, being based upon the literature of the problem.

As indicated above, the learning product, A, is mature when it generalizes to all the situations represented by its complete series, running from $i = 1$ to $i = \infty$. Maturity does not require an infinite number of practices; therefore, the conditions adequate to maturity can be described by the sum, $\sum_{i=1}^{i=n} a_i$. Since the practices are of unequal value, and since the number cannot be definitely determined, adequacy should be conceived as implying an excess, a margin of safety, when maturity is of primary importance, as I represent it as being in connection with the conception of a prerequisite.

The term, a_i, as used above, represents a modified Gestalt. The difference between this concept and that of R. H. Wheeler, for example, can be explained by an analysis of a_i into its constituents. I conceive a_i quantitatively as a product of the elements of order[1] and the elements of detail. Wheeler's whole-minus-the-parts concept (his first organismic law), as one factor, is multiplied by the parts (elements of detail). This makes possible a distinction between a complex-meaningful situation and a simple-meaningful situation which provides the basis for indicating the importance of alternating the highly schematic with the relatively unschemed situations in the practice series. This modification of what appears to be Wheeler's concept can be represented by letting,

$$a_i = \int y_i[(x_i - y_i)],$$

in which

$y_i = Parts$ (Elements of detail),

and,

$x_i = Whole$ (Expressed quantitatively as elements of order plus elements of detail).

The conception of the stimulus as the entire series, corresponding to the matured learning product, incorporates the factor of repetition. There are also certain characteristics revealed by the series as a whole which do not appear when the stimulus is conceived as a single situa-

[1] Density of relations.

tion. One of these, as indicated above, is the possible variation between a_i's in which the form factor, $(x_i - y_i)$, is prominent with those in which the complexity factor, x_i, stands out. This provides for the efficient use of the memory as an aid to the learning process; since an impression of any kind which is intuited in a highly schematic situation can be retained by the memory until matched[1] with its equivalents in a more confused situation. Whatever there is of significance in one situation has to be rediscovered in the other. Also complexities, elements of detail, can be remembered and afterwards joined with the unifying principle in a succeeding, schematic situation. Where cause-and-effect or spatio-temporal relationships compose the pattern of meaning, mathematics (form as expressed through exact, quantitative treatment of specific relationships) defines the extreme limits of the schematic pole.[2]

The axis of a practice series can be defined by: (1) A rational principle; (2) by a goal insight or a series of similar goal insights; or, (3) by a pattern of sensory impressions. The first matures out as a rational insight, the second as a value attitude (a tendency to seek similar goals in the future), and the third is mature when recognition of what is being observed is predominant over new impressions.

When a goal intuition is the objective, literature is the schematic as compared with history and with direct participation in human relationships and situations.[3] According to Herbart, this literature must be the product of genius; I am inclined to agree with this position.

Since the magnitude of the total stimulus-series depends upon both unity and diversity, as either coexisting or as combined through the aid of the memory, a sort of a composite picture of any given stimulus-series can be constructed by superposing all the various members of the series upon each other. If this were possible, the resulting configuration would be too complex for the intuitive powers of any finite learner. Repetition reverses this hypothetical piling up process by thinning out the complexities and making the core intuition more obvious. Thus, repetition is the means of utilizing the memory until

---

[1] See Dewey's suggestive discussion of "matching" in this sense, *Logic, The Theory of Inquiry*, pp. 213–219.

[2] J. A. Lynch, "The Relation of Mathematics to the Learning Process," *Mathematics Teacher*, November, 1939.

[3] J. A. Lynch. "Mathematics in Herbart's Philosophy of Education," *Mathematics Teacher*, January, 1939, which takes account of literature as well as mathematics from Herbart's point of view.

the particular learning product matures. In fact, repetition seems to be necessary in about the inverse ratio to the learner's intuitive powers.

From this point of view, there is no particular reason for assuming that the schematic and the complex should necessarily succeed each other in any mechanical manner. It seems to be sufficient if the two intermingle in some fashion throughout the stimulus-series. This is obvious from the fact that the memory can carry an impression from a schematic to succeeding unschemed situations, or *vice versa*, for an indefinite period of time. From this, it seems evident that the relatively weak memory requires a more frequent shift from one type of situation to the other. Distractions of a kind are not fatal to this process as long as the memory is able to span the time intervals between the practices. In fact, distractions might even prove advantageous to the process in so far as they supply the occasion for needed rest and for spent energies of various kinds, such as physical appetites, to regain their peak strength. There is no good reason that I can see for assuming that the series should begin with one kind of a situation rather than the other, although I am familiar with what both the Gestalt psychologists and the intuitional idealists have to say about beginning with an "insight" or with an "intuition."[1] It seems apparent also that the conditions favorable to an intuition do not always assemble simultaneously; the build-up might extend over an indefinite period of time.[2]

Therefore, priority inside the given stimulus series seems to be indefinite and poorly defined. More important than the order in which the a_i's succeed each other in the series is the assurance that both kinds are included in sufficient numbers to bring about maturity over a wide range of application.

I shall now consider the relationship of priority as existing between one stimulus-series and another. A distinction between the axis and the radius of the stimulus is a convenient approach to this. The part

---

[1] An example of the former is Wheeler's statement of the organismic laws, *The Science of Psychology;* and for a statement of the latter, see William Ernest Hocking's *Types of Philosophy*, pp. 208–212.

[2] An instance is the manner in which my six-year old son explained a battleship he had just made: "This stick looked like a cannon. I needed something to put it on so I found this turret. I sawed the ends off these boards to make a battleship to put the gun on; and I found this stick for a mast. The spool up there is the lookout!" Undoubtedly, he will make another battleship sometime later, and with greater success, if he follows the pattern which I have observed in his previous behavior.

repeated, corresponding to the learning product by which the stimulus-series is defined, can be described as the axis of the stimulus. As indicated already, this can be of any one of the three classes of learnings: (1) Rational intuitions, (2) goal intuitions, or (3) sensory intuitions.[1] Whenever a series of similar goals defines the axis of the stimulus, ideas and other types of learning products become radial in so far as they help clarify (set in relief) the goal insight, the situation which shapes the approach to the goal. If emphasized to the extent of excluding other types of learning situations, this suggests instrumentalism. On the other hand, goals are radial to a rational intuition whenever calling attention to them helps clarify or condition the intuition. Sometimes this process might even require a dramatization of certain involved goals; first, to clarify them and, second, to call attention to their connection with the intuition. This suggests that certain ideas cannot be easily understood out of relation to their so-called practical applications. In this sense, goals simply assume the status of parts (as elements of detail) in relation to the rational intuition around which the configuration is organized. In a similar manner, either goals or ideas might be radial to a pattern of sensory impressions.[2] For instance, a knowledge of the gross relationships between planetary bodies is an aid to the process of making a correct observation of an eclipse. This is due to the fact that the intelligent observer partially deduces from a wider, generalized knowledge and partially apprehends directly a correct impression of the event.

. . . .

My first and most fundamental point is that *the support given to the axial learning by the radial is not of material value until the latter is mature.*

An observer will not know what to look for in an eclipse unless his knowledge of planetary relationships is real. He can, of course, get an accurate sensory impression capable of retention without this support, but it will take longer (require more repetitions). Furthermore, if a rational intuition happens to be functionally related to a certain class of goals, it is possible that the former cannot be clarified readily without connecting it up with some of the so-called practical situations. Where

[1] Implying only the etymological meaning, *intueri*, to look on.

[2] For a good example of a series built up around a particular sensory pattern see Charles De Garmo's three sonnets, "Red," "Blue," and "Yellow," *Aesthetic Education*, pp. 69–71. It is not implied that this is the interpretation put upon these sonnets by De Garmo.

this relationship holds, it is not merely necessary to bring some members of the class of related goals into the various configurations, but there must also be a background of maturity on the part of the corresponding value attitude before the practical situation can be made interesting enough to influence the learning process. In a similar manner, a mature rational intuition might set in relief a related goal, and thereby facilitate the process of maturing the value attitude corresponding to the goal.

An example of the manner in which one rational intuition might condition another can be taken from geometry: The intuition which generalizes out as, "The sum of the interior angles of any triangle equals two right angles," conditions the learning product, "The sum of the interior angles of a polygon equals n minus two times two right angles," where n equals the number of sides. These learnings should be treated as separate quanta in order that proper attention can be given to the maturity of each. The Dewey philosophy, because of its associational bias, encourages the blending and confusion of a new learning product with its prerequisites, as if the whole mass of related ideas could be treated as simply an expanding complex of meaning.[1]

In the immature stage, the learning product is dependent for expression upon objective conditioning; and the memory either completely replaces or supplements the intuitive powers. As mature, the learning product tends to influence the will.[2] It seems obvious that an immature learning product, one which is still dependent upon memory as an aid, can occupy only the central place in a configuration; that is, it can attach to the axis but not to the radius of the stimulus. If this point is well-taken it accounts in part for the weakness of much present-day educational practice, because neither the time nor the occasions for the kind of repetitions which mature learnings to the point at which they effectively condition other learnings are consistently provided.

. . . .

A second important point concerns the *degree of priority in time of one learning product to another.*

If one learning product seems to precede another in the arrangement of the curriculum, how much earlier should the one occur than the

---

[1] There is, however, a tendency toward a similar differentiation of learning products in *Experiences and Nature*, particularly in those chapters dealing with mind and with ideas; Chapters VI, VII, VIII, and IX.

[2] My article, "The Problem of Transfer," deals more fully with this point.

other? An answer to this can be suggested by what might appear out of context to be a dogmatic statement:

In general, where a learning product, A, is a prerequisite to another learning product, B, the larger the number of repetitions required to mature A, the more apparent is the time spread between the initial presentation of A and the initial presentation of B.[1]

This spread of the time interval, which depends upon the average[2] number of practices required, determines the significance of what is often referred to as mental maturity. The relationship between the range of applications of a learning product and its maturity is not easily recognized, because the memory so easily replaces the intuition in connection with highly abstract learning products. For instance, an immature student of mechanics can state the laws of motion long before he is able to philosophize about their complete meaning. Until he can do the latter, however, the learning product will not influence the perspective through which he approaches new learning situations very much if any. An assumption which I have tried before to justify[3] is that the product of the complexity and the unity factors in the stimulus measures the strength of the intuition required to penetrate its meaning. Repetition is represented as not only strengthening the unifying intuition, but as also making possible the distribution of complexities by lengthening the series. Complexity is conceived in terms of both the difficulty of the problems incorporated in the series and the variety and range of situations to which the core intuition is relevant.

The conditions which mature out most so-called philosophical ideas are more complex, in the above sense, than those to which physical principles are relevant, because the former explain relatively undifferentiated conglomerates of human experience. Therefore, most

[1] Repetition takes the place of time.

[2] The present problem deals only with the relationships between learnings as such, as they occur in the organized curriculum, and not with the problems of individual psychology. Therefore, the term average is significant and should be frequently inserted. Just as mortality tables do not help much as a basis of predicting either the cause or the time of a particular death, the values dealt with here do not give much guidance in matters pertaining to individual behavior. However, it is held that precisely in the same sense in which mortality tables are valuable to the insurance business as a whole, so are these averages, to the degree to which they are accurately represented, valuable to the business of education as a whole. This discussion pertains to the problem of organizing but not of administering the curriculum.

[3] "A Quantitative Description of the Stimulus."

philosophical abstractions never get beyond the discussion stage. Such abstractions are empty when they lack maturity. In the first place, they have not been sufficiently discussed, reflected upon, applied, and especially read about in the literature produced by the creative thinkers in that field; and, in the second place, the range of application and interpretation has not been broad enough to include all that is implied by such abstractions. The schematic members of the stimulus-series leading to the maturity of one of these abstractions are derived almost entirely from the literature. No one is capable of getting a clear insight into the most general human problems without the aid of this literature. The ancient substitute for this literature, which is not adequate on the civilized level, is the guidance coming from the mature minds among the living. Organization gives clarity. Those things which are hardest to organize with respect to their meaning have been shaped very gradually by widely scattered individuals who are capable of intuiting meaning where it is not obvious and of recording it in the literature. To the extent that learnings are general in this sense, an abandonment of the literature, in the interest of a direct attack upon the problems, as they are now supposed to confront us, represents a return to barbarism, because the organization which has been slowly won will be quickly lost and confusion will follow. Superstition will fill the void. It takes both the literature and the problems to give the stimulus-series schematic as well as unschemed members.

. . . .

A third and final point has to do with the interchangeability of prerequisites. Since the prerequisites to a particular learning product are given the status of parts, it follows that they are interchangeable to some extent. There is greater flexibility in some cases than in others. Some learnings can be conditioned[1] by a great variety of background learnings; others require pretty definite sequences. It depends upon whether the prior learning is an indispensable or an interchangeable part. Sometimes a prerequisite may be more of the nature of an aid to a new intuition than a necessary conditioner, as in the case of a knowledge of planetary movements in relation to the

[1] Although this term is not used in the sense given to it by *Reflexology*, the emphasis put upon maturity gives it an analogous meaning. An acquired learning product conditions another, according to I. P. Pavlov, when it becomes "reflexogenous"; this corresponds to maturity.

observation of an eclipse. Where this flexibility is conspicuous, the concept of general maturity, meaning an abundance of matured but unspecified learnings, is usually employed to describe the background. Maturity with respect to the particular learnings constituting this undefined background is just as necessary as it is in connection with the more definite prerequisites.

If a prerequisite is more general than the learning product which it conditions, the support given is deductive in the sense that it makes possible at least a partial anticipation of the new intuition. For example, a mature understanding of the atomic hypothesis facilitates an insight into the principle of oxidation which is a special case of the more comprehensive principle. On the other hand, if the prerequisite is less general than the learning which it conditions, the aid given is inductive in so far as it prepares the way for the latter without actually anticipating it. Although it is apparent that there is no particular fixed sequence here, the general trend seems to be upward, from the less to the more general. As far as any particular stimulus-series is concerned, the process is simultaneously inductive and deductive. This is what is implied by the alternation of the schematic and the unschemed practice situations. But deduction in the latter case depends upon memory until maturity takes place. The more teaching is emphasized, the more this use of memory will come into prominence, and it is a mistake to try to identify teaching altogether with the methods used in discovering new truths. The process of discovering new truths is deductive to the extent that a projected hypothesis or intuition controls observation. This is more prominent than is sometimes supposed if one is to judge from the personal histories of such thinkers as Galileo, Ricardo, and Rousseau.[1] While induction and deduction do not seem to determine the relationship between series absolutely, the normal process, as far as the general movement is concerned, seems to be characterized by induction as it is defined above. The problem is obviously a different one within the series from what it is between distinct series. In the paper, "The Problem of Transfer," referred to above, I dealt with this relationship between the more and the less general as if it were definitely fixed in order to set off more clearly the issues involved. The order of priority is not the fundamental problem in that paper; and, although the argument is simplified somewhat by

---

[1] See Rignano, Eugenio: *The Psychology of Reasoning*, Ch. IV, for some interesting examples of reasoning.

dealing with the order as fixed, the conclusions reached do not depend upon it.

CONCLUSIONS

The following generalizations concerning learning readiness are suggested by the foregoing discussion:

1. Prerequisites must be mature.

2. Prerequisites play the rôle of parts in relation to a given learning situation.

3. Prerequisites are not rigidly fixed, because parts are more often than otherwise capable of being replaced by other parts.

4. Any type[1] of matured learning product can function as a part in relation to any other type.

5. The tendency, perhaps only slightly dominant, is for the less general learnings to prepare the way for the more general, as an inductive process;

(*a*) Where a mature, general learning conditions a less general, the support given is deductive in nature;

(*b*) The degree to which the aid given by a prerequisite is deductive, determines the degree to which mental maturity is in evidence, since the more general learnings require more time for maturity where time is represented by repetitions.

6. On the whole, the sequence is less definite from this viewpoint than suggested by the association psychologies and the apperceptive principle, but where it is definite, it must be rigorously observed.

7. Maturity is equally important, whether the background is definite and fixed or general and undefined.

---

[1] Sensory pattern, rational intuition, or goal intuition.

THE VALUE OF INTELLIGENCE, PERSONALITY, AND VOCATIONAL INTEREST TESTS IN A GUIDANCE PROGRAM[1]

LOUISE E. ALTENEDER

State Teachers College, Paterson, N. J.

Various techniques have been developed and used in the field of guidance in colleges, such as the personal interview, observation, biographical records, case studies, cumulative records, and trained personnel staffs including nurse, physician, psychologist, psychiatrist, and personnel worker have been appointed. In addition to these techniques, psychological tests and questionnaires of various kinds are now being used in many institutions.

The present study gives results of four psychological tests which were used as part of a guidance program for college freshmen. The tests were administered to one hundred fifty men and one hundred fifty women. The following tests were used: The Henmon-Nelson Test of Mental Ability for College Students, Form *B*;[2] The Revised Minnesota Paper Form Board Test, Form *AA*;[3] The Adjustment Inventory, by Hugh M. Bell;[4] and the Vocational Interest Blanks for Men and for Women Students, by Edward K. Strong, Jr.[5]

INTELLIGENCE AND ACHIEVEMENT

The scores for the Henmon-Nelson Test of Mental Ability and for The Revised Minnesota Paper Form Board Test show the variation usually found in college freshmen groups. Statistical results are shown in Table I.

The Henmon-Nelson Test norm of 44 was determined on the basis of scores made by approximately fifty-five hundred freshmen in colleges and universities of varying size in different parts of the country.[6]

[1] This is a digest of a thesis submitted for the Ph.D. degree at New York University in 1938.

[2] Henmon, V. A. C. and Nelson, M. J.: *Henmon-Nelson Tests of Mental Ability.* Boston: Houghton-Mifflin Company, 1932.

[3] Quasha, Q. H. and Likert, R.: *The Revised Minnesota Paper Form Board Test.* Psychological Corporation, N. Y.

[4] Bell, Hugh M.: *The Adjustment Inventory.* Stanford University Press, Stanford University, California.

[5] Strong, Edward K., Jr.: *The Vocational Interest Tests for Men and for Women.* Stanford University Press, Stanford University, California.

[6] Henmon, V. A. C. and Nelson, M. J.: *Teacher's Manual for Henmon-Nelson Tests of Mental Ability*, 1932.

The difference between the means for the men and the women in the present study is 4.90, the E_d 1.22, and the ratio 3.84, showing a statistically significant difference. In a recent study by Livesay,[1] using the American Council Psychological Examination with high-school graduates, five hundred men and three hundred seventy-two women, similar results were found. In Livesay's study the women did better

TABLE I.—STATISTICAL RESULTS—INTELLIGENCE TESTS

	Mean	Median	SD
Henmon Nelson Test:			
Men	45.57	46.05	11.80
Women	40.67	41.19	10.25
Men and women	43.30	43.00	11.30
Test norm	.....	44.00	10.20
Minnesota Form Board:			
Men	41.60	42.98	10.05
Women	40.70	42.02	9.95
Men and women	41.15	42.46	10.10
Test norm	.....	38.00	

than the men on the Artificial Language and Opposites Tests, but the men exceeded the women on the total score. The mean for the total score for the men was 148.72, and for the women 143.89. The Henmon-Nelson Test is made up largely of problems involving reasoning and general information.

The Revised Minnesota Paper Form Board Test norm of 38 was based on scores for two hundred forty-seven liberal arts freshmen. The median for three hundred forty-four engineering school freshmen from two universities was 43.[2]

It is interesting to note that in the Paper Form Board Test in the present study the women did as well as the men. One girl scored as high as the boys, which was not true for the results of the Henmon-Nelson Test of Mental Ability.

The coefficient of correlation between the results for the Henmon-Nelson Test and those for the Paper Form Board Test for the three hundred cases is .45 ± .03. This coefficient would indicate that both

[1] Livesay, T. M.: "Sex Differences in Performance on the American Council Psychological Examination." *Journal of Educational Psychology*, Vol. XXVIII, December, 1937, pp. 694–702.

[2] Likert, R. and Quasha, W. H.: *Manual of Directions for The Revised Minnesota Paper Form Board Test.*

similar and different traits are measured by the two tests. Some students were high in one test and low in the other. The question arises as to just what factors of intelligence the Revised Minnesota Paper Form Board measures.

The results for the Henmon-Nelson Test of Mental Ability and those for the Revised Minnesota Paper Form Board Test were correlated with first-semester grades. All grades were used, academic and non-academic. Numerical values from 1 to 5 were assigned to grades *F, D, C, B, A*, respectively. The averages for each student were then obtained and scaled. The results are shown in Table II. The coefficients for both intelligence tests with first-semester grades are positive, those for general ability (Henmon-Nelson Test) being higher than those for mechanical intelligence (Minnesota Paper Form Board). The coefficients for the women are higher than those for the men for both tests. Many studies have been made comparing intelligence and scholarship. Freeman[1] reports that tests of mental ability are positively correlated with actual college marks, the coefficients ranging roughly from .20 to .70.

TABLE II.—INTELLIGENCE AND FIRST-SEMESTER GRADES

	Henmon-Nelson Test		Paper Form Board	
	r	PE	*r*	PE
Grades:				
Men (150)	.35	.05	.29	.05
Women (150)	.44	.04	.33	.05
Men and women (300)	.35	.03	.30	.04

Coefficients of correlation between the Henmon-Nelson Test of Mental Ability and grades for three successive semesters are shown in Table III. An increase is indicated in the coefficients for each successive semester. Studies by Freeman[2] and Darley[3] report a decrease in the coefficients for intelligence and achievement for three successive semesters. The number of cases in both of these studies was larger than the number indicated in Table III. About half of the with-

[1] Freeman, Frank S.: "Predicting Academic Survival." *Journal of Educational Research*, Vol. XXIII, February, 1931, pp. 113–123.

[2] *Ibid.*

[3] Darley, John G.: "Scholastic Achievement and Measured Maladjustment." *Journal of Applied Psychology*, Vol. XXI, October, 1937, pp. 485–493.

drawals in the present study were caused by transfers and half were dropped or left because of poor scholarship. Of those who left or were dropped five fell at or below the tenth percentile on the Henmon-Nelson Test, three at or below the first quartile, and two at or below the median.

TABLE III.—GENERAL INTELLIGENCE AND GRADES FOR THREE SEMESTERS

	Semesters					
	1		2		3	
	r	PE	*r*	PE	*r*	PE
Henmon-Nelson Test.....	.41	.06	.52	.05	.53	.05

N = 83 for semesters 1 and 2; 59 for semester 3

PERSONALITY, INTELLIGENCE, AND ACHIEVEMENT

The Adjustment Inventory for college students by Hugh M. Bell provides four separate measures—home, health, social, and emotional adjustment.[1] The scores for these measures may be combined into a total personality score. Descriptive ratings are provided as norms, such as "excellent," "good," "average," "unsatisfactory," and "very unsatisfactory."

The scores for the Bell Adjustment Inventory in this study tend toward a normal distribution and show individual differences almost as marked as those for intelligence.

Table IV shows results treated statistically, low scores being the best. The findings of this study and also those of Bell, indicate that the boys are better adjusted socially and emotionally, and the girls better adjusted with respect to home and health. The question arises whether the boys are more frank in answering questions relating to home and health, and more reticient regarding social and emotional problems. If this were true it would still show a personality difference.

The coefficients of inter-correlation for the adjustment measures are presented in Table V. The findings of Bell are also given. Bell does not state whether his findings are for men or women.[2]

[1] Bell, Hugh M.: *The Theory and Practice of Student Counseling.* Stanford University Press, 1935.

[2] *Ibid.*

Both the present study and the one by Bell indicate a rather definite tendency for students who are maladjusted toward their home to be maladjusted physically and emotionally. The highest coefficient in the present study is for emotional and health adjustment, and emotional and social adjustment.

TABLE IV.—STATISTICAL RESULTS FOR BELL ADJUSTMENT INVENTORY

	Mean		SD	
	Men	Women	Men	Women
Measure of adjustment:				
Home	7.36	6.87	4.62	5.18
Health	6.60	6.38	4.00	4.04
Social	11.6	13.14	6.66	6.87
Emotional	8.18	12.40	5.31	6.42
Total	32.43	37.37	15.00	16.35

TABLE V.—INTER-CORRELATION OF ADJUSTMENT INVENTORY MEASURES

	Present study				Bell	
	Men		Women			
	r	PE	*r*	PE	*r*	PE
Adjustment measures:						
Home and health	.35	.05	.34	.05	.43	.04
Home and social	.07	.06	.16	.05	.04	.05
Home and emotional	.48	.04	.49	.04	.38	.04
Health and social	.27	.05	.16	.05	.24	.04
Health and emotional	.60	.04	.52	.04	.53	.03
Social and emotional	.40	.05	.42	.05	.47	.04

Correlation coefficients for general intelligence, as measured by the Henmon-Nelson Test; mechanical intelligence, as measured by The Revised Paper Form Board Test; and achievement, as shown by first semester grades, with measures of adjustment on the Bell Adjustment Inventory, are presented in Table VI.

Table VI indicates that the highest positive coefficient is between emotional adjustment and mechanical intelligence, and emotional adjustment and grades for the women. It is interesting to note that the coefficient for health adjustment and grades for the men is −.20

and for the women .15. Social adjustment correlates positively with intelligence and grades for both men and women.

TABLE VI.—INTELLIGENCE, ACHIEVEMENT, AND PERSONALITY[1]

Adjustment measures	Home, *r*	Health, *r*	Social, *r*	Emotional, *r*
Henmon-Nelson:				
Men	−.09	−.07	.14	.09
Women	−.14	−.02	.09	.05
Paper Form Board:				
Men	−.00	−.00	.06	.03
Women	.02	.08	.11	.23
First semester grades:				
Men	−.09	−.20	.17	−.05
Women	−.04	.15	.18	.28
PE of *r*, 0 = ±.06				

[1] In finding the coefficients in Table VI, the adjustment measures were reversed, so that in all the variables the highest score is the best.

Wrenn, and others,[1] found no relationship between level of ability, using the American Council Psychological Examination, and the traits of introversion, extroversion, and emotional stability as measured by the Bernreuter inventory. These authors conclude that

(1) Extremes of intelligence do not appear to be associated with difference in degree of emotional stability.

(2) Highly intelligent junior-college students appear to be more self-sufficient and independent both mentally and socially.

VOCATIONAL INTEREST, INTELLIGENCE, AND ACHIEVEMENT

Table VII presents the number of ratings obtained for the occupations scored on Strong's Vocational Interest Blanks. According to Strong[2] a rating of "*A*" signifies that the individual has the interests of men successful in a particular occupation; "*B* plus," "*B*," and "*B* minus," that the individual probably has the interests; "*C*" signifies that the individual does not have the interests of those successful in a particular occupation.

---

[1] Wrenn, C. Gilbert, Ferguson, Leonard W. and Kennedy, John L.: "Intelligence Level and Personality." *Journal of Social Psychology*, Vol. VII, 1936, pp. 301–307.

[2] Strong, Edward K., Jr.: "The Vocational Interest Test." *Occupations*, Vol. XII, April, 1934, p. 49.

As to the permanence of vocational interests, Strong[1] states that though there are real changes from fifteen to thirty years of age, interest patterns are surprisingly stable from fifteen years on.

TABLE VII.—RATINGS OBTAINED FOR EACH OCCUPATION

Rating	*A*	*B*+	*B*	*B*−	*C*+	*C*
Men:						
Certified public accountant	0	1	6	6	..	137
Doctor	11	13	16	16	..	94
Engineer	11	18	13	4	..	104
Lawyer	5	8	43	23	..	71
Life insurance salesman	2	4	23	16	..	105
Teacher	10	10	25	21	..	84
Women:						
Librarian	3	0	12	18	36	81
Life insurance saleswoman	0	3	2	9	16	120
Teacher	28	28	12	19	16	47
Lawyer	7	10	8	23	16	86
Nurse	57	28	29	15	12	9
Stenographer, secretary	75	40	19	11	2	2
Social worker	5	5	17	13	15	95

The ratings given in Table VII are for inventory interests, regardless of the occupational choices of the students. Fryer states that inventoried interests have been found to be a more valid measure than subjective interests.[2] Table VIII gives the ratings for students choosing the occupations scored.

In making a study of the diagnostic value of the Vocational Interest Test, Strong[3] found that of one hundred fifty-six seniors, forty-six per cent entered the occupation on which they scored highest, twenty per cent the occupation on which they scored next highest, and eleven per cent the occupation on which they scored third highest. Only eighteen per cent entered an occupation for which, according to the test, they had no interest.

Table IX presents coefficients of correlation for vocational interest, intelligence, and achievement. While the coefficients are not statistically significant, some interesting relationships are indicated. The coefficient for the vocational interest for doctor is higher when that

[1] Strong, Edward K., Jr.: "Permanence of Vocational Interests." *Journal of Educational Psychology*, Vol. XXV, May, 1934, p. 336.

[2] Fryer, Douglas: *The Measurement of Interests*. Henry Holt and Company, 1931, p. 203.

[3] Strong, Edward K., Jr.: "Diagnostic Value of the Vocational Interest Test." *The Educational Record*, January, 1929, pp. 59–68.

TABLE VIII.—RATINGS FOR STUDENTS CHOOSING OCCUPATION

Ratings	*A*	*B*+	*B*	*B*−	*C*+	*C*
Men:						
Teacher	8	4	9	8	..	6
Doctor	3	0	3	1	..	3
Lawyer	2	0	1	0	..	0
Engineer	4	3	3	1	..	2
Certified public accountant	0	0	1	1	..	3
Life insurance salesman	0	0	1	0	..	0
Women:						
Librarian	0	0	1	0	0	0
Life insurance saleswoman	0	0	0	0	0	0
Teacher	24	19	9	13	8	21
Lawyer	0	0	1	0	1	0
Nurse	5	0	0	0	0	0
Stenographer secretary	13	3	0	0	0	0
Social worker	0	1	3	0	0	0

vocational interest is correlated with mechanical intelligence than with general intelligence. This is true also for the vocational interest of engineer. Grades show a positive correlation with the vocational interests of certified public accountant, lawyer, and teacher for the men, and with the vocational interest of librarian, teacher, lawyer, and social worker for the women.

In considering the relationship between intelligence and vocational interest Wrenn[1] studied the interests of two groups of junior-college students, one high, and one low in intelligence. He concludes:

> Data all point to the greater consistency, permanence, and suitability of the vocational choices of students high in academic intelligence as measured by the Psychological Examination than students of low intelligence.

Fryer[2] states that figures would indicate that the lower an individual is in mental age the more he is likely to be in error in vocational interests when the error is measured in intelligence requirements of the vocation. This refers to subjective interests, not inventoried interests.

VOCATIONAL INTEREST AND PERSONALITY

Table X gives coefficients of correlation between the scores on Strong's Vocational Interest Blank and the adjustment measures on

[1] Wrenn, C. Gilbert, "Intelligence and Vocational Choices of College Students." *Educational Record*, Vol. XVI, 1935, pp. 217–219.

[2] Fryer, Douglas: *The Measurement of Interests*. Henry Holt and Company, 1931, p. 192.

Bell's Adjustment Inventory. A study of the table shows some interesting trends. The vocational interest of teacher for the women correlates positively for all adjustment measures. This vocational interest for the men shows a negative relationship for all adjustment measures except social adjustment. The highest positive correlation

TABLE IX.—VOCATIONAL INTEREST, INTELLIGENCE, AND ACHIEVEMENT

Vocational interest	Henmon-Nelson, *r*	Minnesota Paper Form Board, *r*	First semester grades, *r*
Men:			
Certified public accountant	.04	−.13	.21
Doctor	−.03	.17	−.11
Engineer	.15	.22	−.19
Lawyer	.01	−.07	.11
Life insurance salesman	−.28	−.19	−.05
Teacher	.08	.07	.30
Women:			
Librarian	.28	.06	.15
Life insurance saleswoman	.10	−.22	−.20
Teacher	.01	.18	.26
Lawyer	.19	.11	.18
Nurse	−.11	.05	−.10
Stenographer secretary	−.18	−.13	−.28
Social worker	.19	−.02	.16
PE of *r*, 0 = ±.06			

for emotional adjustment, .34, is with the vocational interest of teacher for the women. The coefficient for these two measures for the men is −.08. Social adjustment for the women correlates .39 with vocational interest of lawyer, and .38 with that of social worker.

SUMMARY

(1) The results for the four psychological tests used in this study indicate that the students taking the tests are similar to college freshmen in general.

(2) The men did better than the women on the general intelligence test, the difference being statistically significant.

(3) Intelligence scores correlated positively with achievement, and in line with other studies. Coefficients of correlation of intelligence test scores with grades for three semesters show an increase for each semester.

(4) The women show a higher coefficient of correlation between intelligence and achievement than the men.

TABLE X.—VOCATIONAL INTEREST AND ADJUSTMENT MEASURES

Vocational interest	Adjustment measures			
	Home, *r*	Health, *r*	Social, *r*	Emotional, *r*
Men:				
Certified public accountant	−.08	−.05	.01	−.02
Doctor	−.06	.14	−.23	−.04
Engineer	.23	.25	−.07	.22
Lawyer	−.25	−.09	.01	−.13
Life insurance salesman	−.10	−.12	.07	.01
Teacher	−.12	−.14	.15	−.08
Women:				
Librarian	−.23	−.11	−.11	−.09
Life insurance saleswoman	−.06	−.05	.08	−.14
Teacher	.26	.14	.08	.34
Lawyer	−.12	−.03	.39	.09
Nurse	.13	−.01	.03	.02
Stenographer, secretary	.15	.04	−.11	−.08
Social worker	−.06	−.14	.38	−.01
PE of *r*, 0 = ±.06				

(5) The coefficient of correlation of general intelligence with grades is higher than that of mechanical intelligence and grades.

(6) Of ten students who were dropped or left because of poor grades all fall at or below the median for the general intelligence test, and six fall below the median for the mechanical intelligence test.

(7) Some sex differences are indicated in the measures of adjustment.

(8) There is little relationship between the adjustment measures and intelligence and achievement. The highest positive coefficients are for emotional adjustment and mechanical intelligence, and emotional adjustment and grades, for the women. There is a positive correlation between social adjustment and grades for both men and women.

(9) The coefficients of correlation for vocational interests with intelligence, achievement, and adjustment measures, are not statistically significant. Some interesting trends are shown.

CONCLUSIONS

The findings of this study point to the fact that the psychological tests employed are useful from a qualitative rather than a quantitative point of view.

The Henmon-Nelson Test of Mental Ability, or other test of so-called general intelligence, is useful in a guidance program for college freshmen in locating students who are not working up to their ability, and those students who are having difficulty because of lack of ability. Intelligence tests may have some value for the prediction of academic success but should never be used as a single criterion for the exclusion or dropping of students.

The Revised Minnesota Paper Form Board Test is helpful in securing a factor of intelligence other than general intelligence. In some cases it gives a better picture of a student who is not "abstract minded." It helps to explain good achievement for those who receive low scores on other intelligence tests.

The Bell Adjustment Inventory is a practical instrument in the guidance of college freshmen. It deals with everyday situations and uses terms understood by the student. The Inventory serves as an excellent basis for interviewing and in locating students with problems.

It is important that the scores on the Inventory be used qualitatively rather than quantitatively, and that they be interpreted in the light of each individual case. A good or excellent score in one case may be owing to fine emotional stability, while in another to lack of emotional drive. Lack of emotional drive may be caused by poor physical health, temperament, or a temporary emotional problem. In one case an excellent or aggressive social adjustment score may be an asset; in another, a hindrance. Each factor is influenced by other factors, and by the particular situation and individual in question. Because of these and other reasons scores on the Bell Adjustment Inventory should not be used as entrance requirements except in extreme cases. The Inventory is most valuable when used in connection with an orientation course after entrance. When so used the results should be discussed with the student soon after filling in the questionnaire. It is wise not to file the results for future use.

Strong's Vocational Interest Blanks are helpful to the counselor in assisting students to know more about their vocational interests and in selecting a vocation. Relationships between vocational interest and intelligence and personality traits indicate some interesting trends. More research needs to be done along these lines.

STUDIES IN THE PSYCHOLOGY OF MEMORIZING PIANO MUSIC: III. A COMPARISON OF THE WHOLE AND THE PART APPROACH

GRACE RUBIN-RABSON

New York City

So widespread is the championship of the whole as opposed to the part method of learning that another experiment testing the merits of either seems superfluous. Were this championship to be supported by even more evidence than exists, it would still not be conclusive that the same superiority inheres for learning and memorizing material on a keyboard instrument. For in such learning are incorporated some of the aspects of practically all the areas in which the whole-part hypothesis has been tested.

The thirty-odd experiments performed since the beginning of the century deal with every variety of learning: Of motor, maze, nonsense, sense and logical material. Careful analysis by McGeoch[8] reveals, however, that only six of these are statistically reliable and that the conclusions of these six are so varying and derive from such different kinds of subjects, materials, and experimental procedures that no generalization based on them is valid.

McGeogh[10] levels the very serious charge that many investigators have taken for granted an intrinsic superiority in the whole method of learning and have explained away on some experimental basis any superiority revealed by their statistical data in favor of the part method.

Concerning the superiority of the whole procedure in pure motor learning, the work of Pechstein[12,13] is frequently quoted: "The inherent advantages of part learning are mainly the complete utilization of the transfer items and the avoidance of diminishing returns due to the excessive length of the motor problem"[12] . . . "All waste in part learning occurs in the act of connection and is here traceable almost entirely to the influence of place association."[13]

Insofar as keyboard learning is a motor skill, it remains to be seen to what extent these arguments are pertinent. In pure motor learning where each muscle act is directly dependent on the preceding one, any interruption in the sequence should prove economically destructive. In keyboard performance, however, such dependence is more apparent than real, for only within the smaller musical patterns is this linkage imperative.

Using serially unrelated materials, Crafts[4] found that for learning to substitute letters and digits the whole and the part methods were about equally efficient with massed practice, but that the whole method was consistently superior when the practice was distributed.

Similar results were obtained by Wylie[22] in the serial learning of lists of from eight to twenty-four items, but with no decisive differences in retention after twenty-four hours.

Reed[15] observed no reliable difference in the average scores for whole and various kinds of part procedures in learning four stanzas of poetry, and concluded that the efficiency of learning methods is a function of the individual learner.

McGeoch[9] carries this implication further, asserting that for the superior intelligence the whole method is relatively, as well as absolutely, superior.

Only one study in a non-musical area employed materials having some of the psychological attributes of a musical unit, in which small independent patterns become integrated into a larger indestructable configuration. Smith and Powers,[19] evaluating discrete vocabulary and sentence practice in language learning, found the word or part method superior.

Beeby[1] postulates a fundamental difference in the problem when dealing with verbal and motor learning, in that the former is one-dimensional, incorporating only successive constituent elements, while the latter is two-dimensional, involving simultaneous as well as successive elements. In an experiment treating primarily of the transfer of tracing skills perfected in either hand to the execution of both tracings together, he found some transfer as well as some interference, but observed that the interference was considerably reduced when the coördinated tracings were visible to the subject.

This unilateral approach in a simultaneous act of skill is one phase of the part-whole problem and has already been attacked by the present writer.[17] Using complete eight-measure musical units, as in Beeby's study, the successive elements were approached by the whole method and the emphasis was exclusively on the relative values for memorizing of bringing the right and left hands to smooth memorized performance before coördinating them as compared with coördination from the beginning of the learning.

Again, as in Beeby's study, both transfer and interference resulted, but no statistical differences appeared in the two contrasted procedures either in the learning or in the relearning trials. However, when the

experimental material was transcribed after the relearning, a real superiority was apparent for the unilateral approach in producing precise and accurate retention.

Two hypotheses involving the construction of *visible* patterns have been basic in the experiments in this series: First, that recitation (memorizing) at the earliest possible moment in the learning is of the utmost importance in securing stable concepts;[7] second, that the *sight* of the hands on the keyboard is the most permanent image in memorized performance and should be proceeded by the least possible amount of unseen, purely kinaesthetic hand behavior.

Two experiments dealing with the whole-part problem in piano learning antedate the one already cited. Eberly[5] tested the learning of classic, romantic and modern compositions, each nineteen measures in length by the two methods, but treating the learning and memorizing as discrete functions. The two fundamental assumptions that his subjects were equal in ability and his materials equal in difficulty were unsupported in fact; and since there is neither mention of a rotated experimental design nor any indication of reliable statistical differences, his work can unfortunately not be credited with much reliability. In addition, no evidence is presented concerning the lengths of the three parts into which the nineteen measures were divided—whether or not each segment was a complete musical phrase or figure, or whether the division was made arbitrarily at the measure line. On the *musical* partitioning of the whole into parts may hinge the answer to this question.

Brown[3] did not include memorizing, but limited the learning to smooth performance with the music. Using full-length materials, she concluded that the efficiency of the method is a function of the type of the material. Though the experimental materials were equated even to the number of notes, and each was practiced with metronomic control to the split second, nevertheless, this study seems to be vulnerable on the same points as is Eberly's.

THE PROBLEM

To the writer, it seems that the whole-part problem in memorizing piano material will remain insoluble and purely theoretical as long as the word "whole" is taken so literally. Few compositions are only nineteen measures in length or even several times this number, and the conception of an intricate work makes its length in a sense inevitable,

so that the length and the difficulty become part of the same problem.

For flawless performance with the music, the question is enormously simplified. The number of trials depends, of course, on the skill of the pianist and the difficulty of the material. If no technical difficulties exist, there is no particular reason for stopping at any point except at the end.

To the memorizing, however, an extremely vital factor is added. A previous experiment[16] demonstrated conclusively that the keyboard repetitions alone do not make for stable memorizing or retention; that only when these repetitions are preceded by intensive analytical study is the retention dependable. The necessity for preliminary analysis introduces both practical and psychological complications to the whole-part approach. Shall the subject silently analyze the entire composition before doing the memorizing as a whole at the keyboard? How much of it can he retain, and of what use a minute analysis if too much of it becomes confused? To what point in the learning shall he continue the silent study? And is not an intensive analysis itself a "part" approach?

Since the consistent objective in this series of experiments has been to find memorizing procedures which were not only theoretically and experimentally valid but practically applicable as well, the whole-part problem is here reduced to its simplest form; namely, the consideration of the eight-measure period as a whole.

On a purely musical basis it may be argued that the eight-measure period is a complete structure, incorporating within itself a release, an accumulation, and a relaxation of energy sufficiently satisfying to qualify as an elementary song form. It is a large and complex configuration whose smaller patterns exist not only horizontally in either hand, but vertically as well when both hands together produce a harmonic as well as a melodic integration. The largest independent patterns are presented by the material in each hand. That they may be apprehended as independent configurations not only without violence to the total configuration but even to its ultimate advantage was revealed by the study[17] treating of the unilateral and the coördinated approaches.

The patternings of successive elements in either independent configuration present a new problem. Can they too be torn down into smaller patterns without destroying the unity of the total pattern?

Bingham[2] maintains that the aesthetic unity of a melody derives from the experience of completeness: "The unity (then) which makes

the difference between a mere succession of discrete tonal stimuli and a melody, arises not from the tones themselves; it is contributed by act of the listener. When tone follows tone in such a manner that the hearer can react adequately to each, when the response to the successive members of the series is not a series of separate or conflicting acts but rather in each instance only a continuation or further elaboration of an act already going forward, then the tones are not felt as discrete, separate, independent, but as 'related' to each other. And when, finally, the series of tones comes to such a close that what has been a continuous act of response is also brought to definite completion, the balanced muscular resolution gives rise to the feeling of finality, and the series is recognized as a unity, a whole, a melody."

Ortmann[11] later presented melody as an incorruptible totality: "Tones are associated as all sensory data are associated: By similarity, by contiguity. . . . The psychological status of every tone in a melody is determined by its tonal environment, and by its absolute position in the pitch and time series. . . . The psychological status of every new tone in the melody is not a constant: it changes as each new tone of the melody is heard. . . . It is impossible objectively to change a single tone of a melody without changing the psychological status of the adjacent tones, and to a less degree, that of all other tones of the melody falling within the memory span."

This same point of view was reiterated by the Gestaltist, Von Ehrenfels:[21] "The decisive step in the founding of a theory of Gestalt quality was my own assertion: When the memory image of successive notes are present as a simultaneous complex in consciousness, then an idea (Vorstellung) belonging to a new category can arise in consciousness, a unitary idea, which is connected in a manner peculiar to itself with the ideas of the complex of notes involved."

In view of the conformity of opinion concerning the psychological inviolability of melodic unity, the problem in the present experiment may be stated as follows: The separation of the simultaneous configuration into its two component figurations does not interfere with the reconstruction of the total configuration, but only adds clarity to it. If, in addition to maintaining separate the two simultaneous figures, each of these is further disrupted into the small patterns of its successive elements, what hazard will appear in learning, in retention, and in the precision of the total configuration as evidenced by the ability to transcribe the score after relearning?

THE EXPERIMENT*

No subtleties were introduced into the whole and part approaches. The eight-measure unit was learned either as a whole by memorizing the eight measures without interruption; or by learning each four-measure phrase to smooth memorized performance; or by learning each of the two figures in each four-measure phrase before attempting the next. After each phrase had been learned, or all four of the figures, the small parts were welded and learned as a unit. There were, therefore, three methods of learning: (1) As a whole; (2) two large parts or phrases completed by whole learning; (3) four small parts or figures, completed by whole learning.

Subjects.—The nine adult subjects comprised one man and eight women. With a mean of eleven and two-tenth years of piano training (Table I) and fourteen and four-tenth terms of theoretical experience, they were all either professional musicians, or at least skillful pianists. Ranging from nineteen to twenty-six years of age they were fairly homogeneous from the standpoint of optimum learning age[20] but quite diverse when gauged by the range in required learning and relearning trials and in the number of errors made in transcribing the experimental scores.

TABLE I.—THE MEANS AND RANGE OF THE AGE, MUSICAL TRAINING, AND EXPERIMENTAL ACHIEVEMENTS OF THE GROUP OF SUBJECTS AS A WHOLE

	Means	Range
Age	22.4	19–26
Piano study (years)	11.2	7–16
Theoretical training (terms)	14.4	2–24
Learning trials	15.4	9.7–20.5
Relearning trials	8.4	6.1–12.0
Transcription errors	11.8	1.0–27.8

Experimental Compositions.—Nine fairly simple eight-measure units, diverse in style, were adapted from compositions unfamiliar to the subjects. Simplicity was essential because the complete learning of each was accomplished at one sitting. That they differed considerably in difficulty can be seen from Table II, where the most difficult averaged three times the required trials of the easiest.

* The experiment was conducted in the studio of the experimenter in New York City during the Summer of 1938.

The division into parts follows as nearly as possible the natural arrangement of the musical figures. In some cases the division is arbitrary since the figures for right and left hand do not always terminate at the same moment, nor does the hand-pattern necessarily follow the shape of the musical figure. Furthermore, even a figure whose last tone gives some feeling of finality may find its natural tonal resolution in the first tone of the next figure.

TABLE II.—THE VARIATIONS IN DIFFICULTY OF THE EXPERIMENTAL COMPOSITIONS AS SHOWN BY THE AVERAGE LEARNING AND RELEARNING TRIALS AND BY THE AVERAGE ERRORS IN TRANSCRIPTION

Composition number	Learning trials	Relearning trials	Transcription score
1	12.2	5.7	9.4
2	18.3	8.3	12.7
3	16.2	8.2	9.2
4	12.3	7.0	14.4
5	8.6	6.5	8.6
6	12.8	8.2	6.5
7	15.1	8.2	11.4
8	22.6	12.5	14.0
9	20.7	11.2	20.0

To the original whole-part problem, therefore, is added an additional complexity: To discover the effect not only of learning parts which are an integral part of the whole, but also the effect of learning sections whose natural unity has been forcibly disrupted.

Experimental Design.—Fundamental to the use of subjects who differ in learning capacity no matter how carefully they are selected, and to the use of varying experimental materials which can be equated only by the most exhaustive preliminary investigations, is the adoption of an experimental design which assumes these inequalities but controls them. The Latin-Greek Square lends itself admirably to this purpose.

In the present instance, the nine subjects were divided into three groups of three each. Each group performed the experiment three times, learning three compositions by each of three methods, or a total of nine compositions.

The methods, the compositions, and the order of the learning were rotated, so that every combination of method and composition was presented in every order.

Each of the three groups followed a Latin-Greek Square of different construction, and the designs for the three groups in turn constituted a large square. This procedure involves eighty-one learnings, twenty-seven by each of three methods, and so controlled that the varying capacities of the subjects operate equally, as do the varying difficulty of the compositions.

Procedure.—Throughout, only the subject and the experimenter were present. Each keyboard learning was preceded by a seven-minute period of intensive analytical study. The eight measures were seen both as a whole and in small detail, though the subject did not know until the keyboard learning by what method he was to work.

Whether by whole or by part, a routine cycle of right hand alone—left hand alone—both hands together was maintained until one flawless memorized performance was achieved for each. Infractions of note or rhythm were corrected immediately. In the two-part and four-part methods, each fragment was practiced to perfection and in succession, after which the learning, following the same right-left-coördinated cycle, was completed as a whole.

The subject was warned against any rehearsal outside of the laboratory. Exactly two weeks later, relearning was accomplished without preliminary study, as a whole, and without any separate hand practice. After relearning, each composition was transcribed on music paper, from memory, as a further check on differences in the methods.

Because of varying playing speeds, the time for learning was not controlled. Keeping the subject to any fixed metronomic speed embarrasses him and impedes the learning. Each trial was checked.

ANALYSIS OF DATA

As in the previous experiments in this series, two basic statistical techniques measure the significant differences among the methods and estimate the proportional influence of each of the several variables, including the methods themselves.

Learning Trials.—In computing the number of trials required by each method, one right-hand—left-hand sequence was treated as the equivalent of one coördinated trial. In the whole method, therefore, the total learning trials consist of the right- and left-hand trials, divided by two, plus the coördinated trials. In the two-part method, when right-left and coördinated trials have been thus summed, the result is divided by two again to give the equivalent of the number of whole trials, before adding the additional number of trials needed

to weld the parts together. The computations were the same for the four-part method, except that before adding the additional trials needed to weld the four parts together, the summed result of the parts was divided by four.

The learning trials are presented in two ways: First, according to the number of *right*, *left*, and *coördinated* trials required by each method, and, second, according to the *total* number of trials required by each method.

From Table III it can be seen that there is little difference in the means of the three methods, the right hand showing the largest differences, .61 trials, between the whole and the two-part method. Consistently, however, the fluctuation in the group decreases progressively in the two- and four-part methods.

Table III shows that the slight differences are statistically unreliable, the ratio of the standard error of the difference to the difference of the means being in every case less than one.

TABLE III.—MEANS AND STANDARD DEVIATIONS AND THE SIGNIFICANCE OF THE DIFFERENCES OF THE MEANS OF THE RIGHT-HAND, LEFT-HAND AND COÖRDINATED LEARNING TRIALS BY THREE METHODS

Method	Mean	SD	Difference of means	SE difference	Ratio
Right Hand					
Whole	5.48	2.82	Whole—two parts −.61	.76	.80
Two-part	6.09	2.68	Whole—four parts −.46	.74	.62
Four-part	5.94	2.54	Two parts—four parts −.15	.72	.21
Left Hand					
Whole	7.19	3.78	Whole—two parts −.12	.93	.13
Two-part	7.07	3.04	Whole—four parts −.12	.88	.14
Four-part	7.07	2.44	Two parts—four parts −.00	.76	.00
Coördinated					
Whole	8.85	4.24	Whole—two parts −.30	1.12	.27
Two-part	9.15	3.94	Whole—four parts −.02	.97	.02
Four-part	8.87	2.78	Two parts—four parts −.28	.92	.30

The total learning trial means (Table IV) show practically no differences, but again the standard deviation of the two- and four-part methods are appreciably smaller. The capacities of the subjects show successively less variation when learning in smaller units.

That these differences are negligible appears in Table IV. The ratios are insignificant.

TABLE IV.—MEANS AND STANDARD DEVIATIONS OF THE LEARNING TRIALS AND THE SIGNIFICANCE OF THE DIFFERENCES OF THE MEANS OF THE TOTAL NUMBER OF LEARNING TRIALS BY THREE METHODS

Methods	Mean	SD	Difference of means	SE of difference	Ratio
Whole.....	15.19	9.9	Whole—two parts −.54	2.6	.21
Two-part.....	15.73	9.0	Whole—four parts −.15	2.3	.17
Four-part.....	15.34	6.3	Two parts—four parts −.39	2.1	.19

Since the methods themselves show no differences, to what other factors is the variation among the eighty-one learning scores ascribable? Four variables have been controlled; namely, the subjects, the methods, the compositions, and the order of learning. When these have been estimated, some part of the total variation remains which is due to unknown, hence uncontrollable, factors. In the analysis of the variance, Table V shows this residual error to be only two and nine-tenths per cent, a satisfactorily small proportion. The subjects, accounting for thirty-two per cent, and the compositions, accounting for fifty-five per cent, are here the two significant variables. Method, with only seven-tenths per cent, supports the insignificance of the differences of the means, while order, nine and four-tenths per cent, reveals a considerable adaptation to the experimental situation. Only the ratios for subjects and for compositions exceed Fisher's[6] one per cent value, that value which might be exceeded in random sampling from a homogeneous population once in a hundred trials.

Relearning Trials.—The amount of retention achieved by the various methods after an elapsed time of two weeks can be measured by the number of trials needed for relearning. Relearning by the whole approach might be expected to throw some advantage to the compositions originally learned by this method. But for the relearning trials, (Table VI) the differences in the means are even smaller than in the

learning trials, the largest being forty-nine hundredth trials. Again a smaller deviation derives from the part approaches, proving an increasingly consistent group homogeneity for these routines.

These means naturally show no significant differences. (Table VI.)

The analysis of the relearning variance (Table V) bears this out. Methods produce only one and nine-tenths per cent of the total, while again the subjects and the compositions sustain the major share with forty-three per cent each. The residual error is only four per cent and order eight and one-tenth per cent. As before, only the ratios for the subjects and for the compositions exceed Fisher's values.

TABLE V.—ANALYSIS OF THE VARIANCE OF THE LEARNING AND RELEARNING TRIALS AND OF THE TRANSCRIPTION ERRORS, SHOWING THE RATIO OF EACH VARIABLE TO THE RESIDUAL ERROR, FISHER'S ONE PER CENT VALUE, AND THE PERCENTAGE OF THE WHOLE DUE TO EACH VARIABLE

Variable	Fisher's one per cent value	Learning trials		Relearning trials		Transcription errors	
		Ratio	Per cent	Ratio	Per cent	Ratio	Per cent
Subjects	2.82	11.3	32.0	12.3	43.0	10.9	73.0
Methods	4.98	0.1	0.7	0.43	1.9	0.02	0.6
Compositions	2.82	19.2	55.0	12.3	43.0	2.0	13.4
Order	4.98	3.3	9.4	2.3	8.1	0.94	6.3
Residual error	....	....	2.9	.....	4.0	.....	6.7

TABLE VI.—MEANS AND STANDARD DEVIATIONS OF THE RELEARNING TRIALS AND THE SIGNIFICANCE OF THE DIFFERENCES OF THE MEANS

Method	Mean	SD	Difference of the means	SE difference	Ratio
Whole	8.48	4.2	Whole—two parts .19	1.0	.19
Two-part	8.67	3.4	Whole—four parts .30	.9	.33
Four-part	8.18	2.2	Two parts—four parts .49	.8	.61

Transcription Score.—Previous experiments have indicated that comparable keyboard fluency in memorized performance may obscure differences in clarity and precision in recall. The stability of the learning concepts can be better measured when the overt kinaesthetic func-

tions have been eliminated. The vividness of the inner tonal and visual images acquired through efficient learning insure accurate transcription of the musical score.

Should any ineconomies inhere in the part methods, they might be expected to appear in the errors in the transcription at the points of division, since "all waste in part learning occurs in the act of connection."[13]

TABLE VII.—MEANS AND STANDARD DEVIATIONS OF THE TRANSCRIPTION ERRORS AND THE SIGNIFICANCE OF THE DIFFERENCES OF THE MEANS

Method	Mean	SD	Difference of the means	SE difference	Ratio
Whole.....	11.41	12.0	Whole—two parts .44	3.4	.13
Two-part.....	11.85	13.4	Whole—four parts .85	3.1	.27
Four-part.....	12.26	10.4	Two parts—four parts .41	3.3	.12

Table VII shows the means of the transcription errors and the significance of their differences. Consistent with the learning and relearning comparisons, no differences appear in the means, and though no greater precision derives from one method than another, the deviations among the scores dependent on the part methods are progressively smaller.

Most of the fluctuation among the eighty-one transcription scores is due to the varying capacities of the subjects (Table V) where seventy-three per cent of the total is ascribable to this one variable. The diverse difficulty of the compositions accounts for only thirteen and four-tenths per cent. The ability to transcribe music from memory, though to some extent dependent on individual talent, is, nevertheless, usually the result of specific training.* When this capacity is well

* This does not mean, however, that a high correlation exists between the ability to transcribe accurately and the amount of such training. In this, as in two other experiments where the transcription technique has been used as a check, there is little consistency between the rankings of the subjects by transcription score and by theoretical training. The same inconsistency is apparent in the rankings of piano training and learning speed, and of theoretical training and learning speed. A very constant factor, however, is the high consistency in the rankings of learning and relearning speed and of these and the transcription scores. In the present instance, the correlation between learning and relearning speed is .72, between learning speed and transcription score is .85, and between relearning speed and transcription score, .79.

developed, the added increment of difficulty does not operate significantly. In another experiment[18] using the same subjects and more difficult material, this conclusion was statistically supported.

When evaluated by three separate criteria, the whole method of learning an eight-measure musical score evinces no superiority over part methods. On the contrary, a consistently smaller group deviation for the part approaches furnishes some ground for recommending the latter to a heterogeneous group of music students.

No illogic inheres in the assumption that if violence is not done the economy of the learning of such a tightly-knit unit as the eight-measure period, then none can be expected in learning larger structures in parts, where the relationships of design are less closely integrated.

These conclusions add to the accumulating evidence that the theory of learning by apperception of entire configurations promulgated by Gestaltists may require some modification. Some of this evidence follows:

It may be taken for granted that in learning keyboard music tonal and kinaesthetic patterns must emerge from the background of note sequences before dependable memorizing occurs. These patterns may be intrinsic in the material or appear spontaneously through an independent organization of note groupings by the individual learner. In a study dealing with the merits of massed and distributed practice, definite superiority was manifested for the distributed practice, but the advantage was specific for the less gifted subjects. The following hypothesis was advanced by the writer: "Inherent in musical structure is an organization and grouping of patterns usually independent of the demarcations of the measure; these patterns depend in part on the readiness with which certain keyboard units fall into hand groupings and in part on a grasp of intrinsic small patterns within the large patterns which may not be immediately discernible. Speed in learning and memorizing is directly related to the early perception of these groupings and patterns. For the less able learner, on first acquaintance with the whole configuration, these patterns may not emerge from their background, but must be 're-seen' in another situation after some familiarity with the material has been achieved. During the massed learning, no opportunity is present for this 're-seeing' so that the pattern is never clearly seen against its background and appears only in its first configuration. Such learning is, in a sense, dependent on repetition and not on a logical reconstruction of the units of structure. It can, therefore, produce only a relative amount of retention."

In the study[18] just cited, the learning of much more difficult material was accomplished as a whole, so that the subject was obliged to discover the fairly intricate patterns without help. In the present instance, not only were the patterns in the easier material more comprehensible, but they were marked off for the learning by the experimenter, hence they were, for the most part, perceived at once even by the less gifted subjects.

The advantage of breaking down the simultaneous configuration into its two component figures has already been indicated.[17]

Razran[14] from a study of Gestalt organization and configural conditioning concludes: "There are no definite, specific, static, final laws of pattern organization. Whatever may make for organization at one period or in one situation may cause disorganization at another period or situation, and properties seemingly preventing organization may themselves in time become organizing principles." . . . "The only more or less defined principle of patternization—or organization—is the simple-complex trend. . . . Just what is simple and what is complex, the distance and steps between them, are specific problems to be decided by specific experiments, in which the level of responding, the developmental stage of the subjects, the nature of the task, and the subjects' familiarity with the task, are no doubt the most significant variables."

SUMMARY AND CONCLUSIONS

Contrary to popular and even scientific belief, there is insufficient reliable experimental support for the general assumption that whole learning is in some way intrinsically superior to part learning.

Whatever the ultimate conclusions concerning this superiority may be for motor, maze, nonsense, sense and logical material, no inferences from them can be applied to such an area as keyboard-music learning since the factors which make the latter unique tend to obscure the factors which make it in some ways similar to the others.

For experimental purposes, some practical definition of a musical "whole" becomes imperative. In view of the great length and difficulty of much of the piano literature, conclusions derived from the use of complete musical compositions of arbitrary length must perforce remain theoretical, or at least practically inapplicable for most music students. If the problem is one not only of learning, but of memorizing as well, then the added factor of immediate keyboard memorizing

preceded by intensive analytical study automatically eliminates the use of larger forms.

The eight-measure period is the smallest complete musical form and provides the basic structural unit for all larger forms. The small patterns of which it is constructed are more intrinsically related and more closely integrated than are any multiples of which it may in turn serve as the small pattern. It is, therefore, reasonable to assume that if any ineconomies inhere in the part methods they will be most readily apparent in a form in which the inner relationships of the parts constitutes a psychological unity.

Nine units of approximately eight measures each were adapted from various styles of unfamiliar piano material, and learned either as a whole, in two phrases, or in four small figures. The divisions followed the musical structure as closely as possible, but in order to secure termination of the figure in both the right and the left hand at the same moment, an occasional disruption of the hand pattern or a severing of a melodic tone from its natural resolution was unavoidable. These contradictions contributed an extra complicating hazard to the part approach.

Nine adult subjects, all skillful pianists, repeated the experiment three times, memorizing three compositions by each of three methods. The several variables were presented in every possible combination and order, so that the twenty-seven learnings by each method can be reliably compared.

Each keyboard memorizing was preceded by an analytical study period of fixed length. All subsequent keyboard procedure followed a cycle of one right hand—one left hand—one coördinated trial, until the criterion of flawless memorized performance was achieved.

The relearning to measure retention took place two weeks later, without study period or separate hand practice. After each memorized relearning, the subject transcribed the score on music paper as a further check on the accuracy of the retention.

Neither the means of the methods of the learning nor of the relearning trials show any differences. No superiority can be found in either of the three approaches, nor is retention in any way affected by having first learned the whole in smaller parts. Since the keyboard fluency occasionally obscures differences in the clarity of the details of the material in recall, the transcriptions might be expected to show some weaknesses at the points of juncture of the smaller parts. But no differences exist in the means of the transcription scores.

Consistently, however, the standard deviation of the group decreases progressively as the units of learning are smaller, so that there is some ground for believing that a group of music students of diverse capabilities could most profitably be advised to work in as small units as they individually find satisfactory.

On the other hand, the larger standard deviation for the whole approach indicates that the ability to work comfortably with larger units is specific to the more capable subjects and so becomes a criterion toward which to strive. Furthermore, the musical value of thinking in larger units, quite apart from any consideration of learning efficiency, recommends that the length of the unit be increased as the student grows in capacity.

Despite the violence done to the psychological unity of the eight-measure fragment by tearing it into small parts, no evidence of this appears in the statistical results. There is reason to believe that the Gestalt theory of irreducible configurations may be faulty and that the smaller elements of the whole may not only be perceived as such, but may be welded successively into a whole by a simple-to-complex progression.

BIBLIOGRAPHY

1. Beeby, C. E.: "An experimental investigation into the simultaneous constituents in an act of skill." *Brit. J. Psych.*, Vol. XX, 1930, pp. 336–353.
2. Bingham, W. Van Dyke: "Studies in Melody." *Psych. Monog.*, Vol. XII, No. 3, whole No. 50, 1910.
3. Brown, R. W.: "A comparative study of the whole, part, and combination methods of learning piano music." *J. Exper. Psych.*, Vol. XI, 1928, pp. 235–248.
4. Crafts, L. W.: "Whole and part methods with unrelated reactions." *Amer. J. Psych.*, Vol. XLII, 1930, pp. 591–601.
5. Eberly, L. E.: *Part versus whole method in memorizing piano music.* Unpublished Master's Thesis, Columbia Univ., 1921.
6. Fisher, R. A.: *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh, 1934.
7. Gates, A. I.: "Recitation as a factor in memorizing." *Arch. Psych.*, No. 40, 1917.
8. McGeoch, G. O.: "Whole-part problem." *Psych. Bull.*, Vol. XXVIII, 1931, pp. 713–739.
9. McGeoch, G. O.: "Intelligence quotient as a factor in the whole-part problem." *J. Exper. Psych.*, Vol. XIV, 1931, pp. 333–358.
10. McGeogh, J. A.: "Acquisition of skill." *Psych. Bull.*, Vol. XXVIII, 1931, p. 418.
11. Ortmann, O.: "On the melodic relativity of tones." *Psych. Monog.*, No. 162, 1926.

12. Pechstein, L. A.: "Whole or part methods in motor learning. A comparative study." *Psych. Monog.*, No. 99, 1917.
13. Pechstein, L. A.: "Alleged elements of waste in learning a motor problem by the part method." *J. Educ. Psych.*, Vol. VIII, 1917, pp. 303–310.
14. Razran, G. H. S.: "Studies in configural conditioning: IV. Gestalt organization and configural conditioning." *Jour. Psych.*, Vol. VII, 1939, pp. 3–16.
15. Reed, H. B.: "Part and whole methods of learning." *J. Educ. Psych.*, Vol. XV, pp. 107–115.
16. Rubin-Rabson, G.: "Influence of analytical pre-study in memorizing piano music." *Arch. Psych.*, Col. Univ., No. 220, 1937.
17. Rubin-Rabson, G.: "Studies in the psychology of memorizing piano music: I. A comparison of the unilateral and the coördinated approach." *J. Educ. Psych.*, Vol. XXX, 1939, pp. 321–345.
18. Rubin-Rabson, G.: "Studies in the psychology of memorizing piano music: II. A comparison of massed and distributed practice." *J. Educ. Psych.*, Vol. XXXI, 1940, pp. 270–284.
19. Smith, S. and Powers, F.: "Relative value of vocabulary and sentence practice for language learning." *J. Soc. Psych.*, Vol. I, 1930, pp. 451–456.
20. Thorndike, E. L., Bregman, E. P., Tilton, J. W., and Woodyard, E.: *Adult Learning.* Macmillan, New York, 1928.
21. Von Ehrenfels, C.: "On Gestalt qualities." *Psych. Rev.*, Nov. 1937, pp. 521–524.
22. Wylie, G. E.: *Whole versus part method of learning as dependent on practice.* Doctors Dissertation 1928, University of Chicago.

BOOK REVIEWS

RAYMOND HOLDER WHEELER. *The Science of Psychology.* New York. N. Y.: Crowell & Co., 1940, pp. 436.

This book is a "rewritten, reorganized and abridged" edition of *The Science of Psychology* (1929). That the point of view is "configurational" goes without saying. The author's style is exuberant and enthusiastic—as if he were preternaturally anxious for the reader to get his point of view. This zeal results in an almost complete lack of criticism of researches supporting the Gestalt position. The general spirit of the book is advocatory rather than evaluative.

To the reviewer, the emphasis upon "wholes" is a good one, but when the distinction between "parts" and "wholes" is made a fighting issue, it is often difficult to follow the antagonists. There are not many scientists who take the extreme "mechanistic" position such as Wheeler attacks, and few "organismic" enthusiasts push their beliefs to a logical conclusion. If they did there would be nothing but the "great whole" and that is a bit comprehensive for any mundane mind. Both analysis and synthesis constitute valid approaches to understanding. Parts are always, it is true, parts of something, but they still can be considered as parts and are deserving of scrutiny as such. The textbooks written from the Gestalt point of view are broken down into chapters, or parts, like any other texts.

This argument regarding Gestalten is specious and illustrates Wheeler's preternatural concern over priority of one sort or another:

> If parts existed in the beginning they must have been parts of something at the time they originally existed. Parts of what? What are parts *of* anyway? Wholes, of course. If, then, by the time parts or elements existed in nature, they were already parts or elements of something, then that something must, after all, have been first. It must have been primary. The whole then, is primary, not the parts. (p. 18)

In the course of developing his thesis, Wheeler cites many laws. Almost any defensible generalization becomes a law. Here is an alphabetical list:

1. Law of Action and Reaction.
2. Law of Configuration.
3. Law of Derived Properties.
4. Law of Determined Activity.
5. Law of Field Genesis.

6. Law of Field Properties.
7. Law of Increasing Energy.
8. Law of Individuation.
9. Law of Least Action.
10. Law of Maximum Work.

For many of these laws there is a complementary law to take care of the exceptions. These principles are illustrated less frequently than formerly by reference to physics or chemistry.

Because of his interest in the history of science and in establishing priority, the author gets the cart before the horse every now and then in his discussion of natural laws. They are to him things we assume to make predictions. Such an assumption might well be termed a hunch or a hypothesis, but a law, the reviewer holds, is of sterner stuff. It is well beyond the assumption stage. It is tested by a great number of predictions, many of which are undertaken for the very purpose of checking the law.

Wheeler seems to have developed a conceptual blindness to shades of gray. Almost everything appears to him to be either white or black—all of this and none of that. His assertions lack only the contentiousness of dogmatism. Here are some attributes of the "crowd" as cited in a quotation by the author which illustrate the criticism (p. 46–49).

1. There are *no* problems for the crowd because they are all settled in advance (by beliefs and prejudice).
2. The crowd is *utterly* devoid of humor.
3. The individuals in a crowd are divested of *all* sense of personal responsibility.
4. The crowd *never* thinks in terms of abstract ideas.
5. A crowd is *utterly* unaware that its attitudes exist only to justify self-feeling.

If the author refuses to call any gathering a crowd unless these extreme characteristics are evidenced that is of course his prerogative. The experience of the reviewer with what he has always considered "crowds"—even mobs—has led to different inferences.

The reviewer anticipated most stimulus from the section on learning, but he was disappointed. The author defined learning as "that behavior in terms of which the individual, through time, extends his insight into a given situation and increases the complexity of his actions with respect to a certain goal." (p. 209) To understand this, a definition of insight as "organized response at the level of conscious

behavior" (p. 209) was added. This left the reviewer vague as to consciousness, but he learned it to be "a field property of sufficiently complex organisms in a waking state." (p. 26) The matter was pursued no further.

The present text lacks the profundity and originality of *Principles of Mental Development* (1932) by the same author. That book was thought provoking and stimulating and it was apparent that Wheeler himself was reasoning carefully. The same impression is not given by *The Science of Psychology*—probably because it is intended as an elementary text.

STEPHEN M. COREY.
University of Wisconsin.

CLARK L. HULL, CARL I. HOVLAND, ROBERT T. ROSS, MARSHALL HALL, DONALD T. PERKINS, and FREDERIC B. FITCH. *Mathematico-Deductive Theory of Rote Learning.* New Haven: Yale University Press, 1940, pp. 329.

This work which, according to its subtitle, is "a study in scientific methodology," is a coöperative venture including two psychologists (Hall and Hovland), three mathematicians (Ross, Hall and Perkins), and a symbolic logician (Fitch). The task set by the senior author was to develop a logico-mathematical system of rote learning of nonsense syllables, a very common topic in experimental psychology which lends itself nicely to quantitative handling.

The base of the system is a series of eighteen postulates, with ten corollaries distributed among seven of them. There are postulated stimulus traces resulting from syllable exposure, and excitatory and inhibitory potentials which "summate by simple subtraction." The eighteen postulates are concerned with the relations among these three. In addition to the postulates there are sixteen undefined and eighty-six defined concepts. From these bases fifty-four theorems (frequently with one or more corrolaries) are proved mathematically. Experimental evidence related to the theorem is discussed in respect to the theory. In some cases experimental evidence related to the theorem was not available, in others the experimental facts were in more or less disagreement with the theorems. The most serious disagreement is, according to Hull, with Jost's law. However, in most instances experimental results from the literature agreed with the theory, or experiments planned to demonstrate the truth or falsity of a theorem conformed to it.

This book is not offered as a final systematization of rote learning but as a more complete exposition of the system introduced by Hull in 1935 and since elaborated by the senior author and his students. No one would deny the desirability of logically developed theory as the foundation for research. There may well be serious question whether the elaborate mathematical and symbolic logic methods used in the present endeavor add much to lucidity. That the system is logically consistent with its postulates there is no question; whether the postulates are acceptable is another matter. In so far as this book aims at demonstrating the utility of rigorous mathematico-deductive methods in science (psychology) it has done its work well—but such utility would seem to be obvious. However, it is perhaps unwise, with an admitted incomplete work before us, to evaluate either the system or the method.

C. M. Louttit.
Indiana University.

Marian Milner. *The Human Problems in Schools.* London: Methuen & Co. 1938, pp. 320.

This book embodies the results of an investigation carried out in England by Mrs. Milner, as investigating psychologist, on behalf of the governing body and head mistresses of the Girls' Public Day School Trust. The aim of the investigation, which was originally planned to occupy a year but was subsequently extended to three years, was to study some possible applications of psychological methods in schools.

Section 1 of the book describes the problems Mrs. Milner found, as seen by the staffs and as stated by the girls. Section 2, the longest in the book, deals in some detail with the psychological techniques of observation and measurement used in the experiment and the results obtained by these methods. On the basis of those results Mrs. Milner then discusses, in Sections 3 and 4, the function of the psychologist within the school and practical implications of the investigation.

This book is written in an easy style and has many illustrating examples. It will be of special interest to educational psychologists, but other people in the field of education will find it well worth reading.

Forrest H. Kirkpatrick.
Bethany College.